



STIC Search Report

Biotech-Chem Library

Rush

STIC Database Tracking Number: 120029

TO: Cynthia Wilder
Location: REM-2A35/2C18
Art Unit: 1637
Sunday, April 25, 2004

Case Serial Number: 10/084555

From: Mary Jane Ruhl
Location: Biotech-Chem Library
Remsen 1-B55
Phone: 571-272-2524

maryjane.ruhl@uspto.gov

Search Notes

Examiner Wilder,

Here are the results for your recent search request.

Please feel free to contact me if you have any questions about these results.

Thank you for using STIC services. We appreciate the opportunity to serve you.

Sincerely,

Mary Jane Ruhl
Technical Information Specialist
STIC
CM-1, Rm. 6-A-06
605-1155

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2004, 17:02:49 ; Search time 621.704 Seconds
(without alignments)
1464.047 Million cell updates/sec

Title: US-10-084-555A-115
Perfect score: 21
Sequence: 1 ttgtgtggggagttattgagt 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb_ba.*
- 2: gb_hg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pin.*
- 35: em_htg_rod.*
- 36: em_htg_man.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	18.4	87.6	637	11	BV019251	BV019251 S212P6011
2	18.4	87.6	153955	10	AL6111983	AL6111983 Mouse DNA
3	18.4	87.6	156008	9	AL353033	AL353033 Human DNA
4	18.4	87.6	156227	9	AC073258	AC073258 Homo sapi
c 5	18.4	87.6	183251	2	AC013555	AC013555 Homo sapi
c 6	18.4	87.6	201027	9	AC025105	AC025105 Homo sapi
c 7	18.4	87.6	201725	2	AL591706	AL591706 Homo sapi
c 8	18.4	87.6	223768	2	AC107502	AC107502 Rattus no
c 9	18.4	87.6	260335	2	AC115140	AC115140 Rattus no
c 10	17.8	84.8	12781	6	AX822420	AX822420 Sequence
c 11	17.8	84.8	12781	6	AX826050	AX826050 Sequence
c 12	17.8	84.8	19653	6	AX346237	AX346237 Sequence
c 13	17.8	84.8	110000	2	AL732359	AL732359 Homo sapi
c 14	17.8	84.8	156598	5	BX005055	BX005055 Zebrafish
c 15	17.8	84.8	157494	10	AL606512	AL606512 Mouse DNA
c 16	17.8	84.8	161477	9	AC023829	AC023829 Homo sapi
c 17	17.8	84.8	162821	2	AC018844	AC018844 Homo sapi
c 18	17.8	84.8	172723	10	AL808018	AL808018 Mouse DNA
c 19	17.8	84.8	196139	2	AC128279	AC128279 Rattus no
c 20	17.8	84.8	201114	10	AL808132	AL808132 Mouse DNA
c 21	17.8	84.8	216225	2	AC117605	AC117605 Mus muscu
c 22	17.8	84.8	218573	10	AC111460	AC111460 Rattus no
c 23	17.8	84.8	222485	2	AC118014	AC118014 Mus muscu
c 24	17.8	84.8	243362	2	AC096234	AC096234 Rattus no
c 25	17.4	82.9	1057	9	AH006998S04	AH006998 Homo sapi
c 26	17.4	82.9	4953	8	AF098645	AF098645 Magnaport
c 27	17.4	82.9	93714	2	AL161661	AL161661 Homo sapi
c 28	17.4	82.9	143255	2	AC015996	AC015996 Homo sapi
c 29	17.4	82.9	147815	9	HS938H6	HS938H6 Human DNA
c 30	17.4	82.9	161869	10	AC131757	AC131757 Mus muscu
c 31	17.4	82.9	181152	2	AC110440	AC110440 Rattus no
c 32	17.4	82.9	239603	2	AC132780	AC132780 Rattus no
c 33	17.4	82.9	252075	2	AC095551	AC095551 Rattus no
c 34	17.4	82.9	252477	2	AC136417	AC136417 Rattus no
c 35	17.4	81.0	400	11	G29283	G29283 human STS S
c 36	17.4	81.0	30616	2	AC018027	AC018027 Drosophil
c 37	17.4	81.0	160000	2	AC007140	AC007140 Homo sapi
c 38	17.4	81.0	171226	3	AC008144	AC008144 Drosophil
c 39	17.4	81.0	175781	3	AC008091	AC008091 Drosophil
c 40	17.4	81.0	178251	3	AE003737	AE003737 Drosophil
c 41	17.4	81.0	182406	9	AC079239	AC079239 Homo sapi
c 42	17.4	81.0	188032	2	AC016393	AC016393 Homo sapi
c 43	17.4	81.0	191904	9	AC005699	AC005699 Homo sapi
c 44	17.4	81.0	203225	5	BX255942	BX255942 Zebrafish
c 45	16.8	80.0	813	11	BV036806	BV036806 S212P6018

ALIGNMENTS

RESULT 1
BV019251
LOCUS BV019251
DEFINITION S212P60110FC1 TO CZECHII/Bi Mus musculus STS genomic, sequence
tagged site.
ACCESSION BV019251
VERSION BV019251.1 GI:31103146
KEYWORDS STS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (Bases 1 to 637)
REFERENCE Wade, C.M., Kulbokas, E.J. III, Kirby, A.W., Zody, M.C., Mullikin, J.C.,
Lander, E.S., Lindblad-Toh, K. and Daly, M.J.

TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

The mosaic structure of variation in the laboratory mouse genome
Nature 420 (5915), 574-578 (2002)
22354584
12466852

Contact: Kerstin Lindblad-Toh
Whitehead Institute for Biomedical Research, Center for Genome
Research
320 Charles Street, Cambridge, MA 02141, USA
Tel: 6172521477
Fax: 6172580903
Email: kersli@genome.wi.mit.edu
Primer A: None
Primer B: None
STS size: 637
Protocol:

WGS-discovery: Paired-end low-coverage whole genome shotgun reads
were generated from 129S1/SvImJ, C3H/HeJ, and BALB/cByJ. The WGS
reads were placed uniquely on the WGSVC3 C57BL/6J assembly and SNP
detection was carried out by SSAHA-SNP. 225,000 reads were
annotated
as STSs and 81,000 SNPs were annotated with alleles from C57BL/6J
and the strain from which the particular read came. The validation
rate for these SNPs was estimated at approximately 98%.

FEATURES
source

1. .637
Location/Qualifiers
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/map="4 22-587 126237344-126237908"
/clone_lib="CZCHII/E1"
<1...>637

STS
ORIGIN

Query Match 87.6%; Score 18.4; DE 11; Length 637;
Best Local Similarity 95.0%; Pred. No. 7e-02; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TGTGCGGGAGTTATTGAGT 21
Db 367 TGTGCGGGAGTTATTGAGT 366

RESULT 2
AL611983

LOCUS Mouse DNA sequence from clone RP23-467J23 on chromosome 4, complete
DEFINITION sequence.
AL611983
VERSION AL611983.23 GI:29373965
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Clark, G.

Direct Submission
Submitted (30-MAR-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk

On Mar 30, 2003 this sequence version replaced gi:28300674.
Sequence from the Mouse Genome Sequencing Consortium whole genome
shotgun may have been used to confirm this sequence. Sequence data
from the whole genome shotgun alone has only been used where it has
a phred quality of at least 30.

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT
Submitted (07-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk
On Nov 15, 2001 this sequence version replaced gi:15131975.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30);
an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30);
an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information
on the WormPEP database can be found at
http://www.sanger.ac.uk/projects/c_elegans/wormpep RP23-467J23 is
from the RPCI-23 Mouse BAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6.

FEATURES
source Location/Qualifiers

1. .153955
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="4"
/clone_lib="RPCI-23"
/clone_lib="RPCI-23"

ORIGIN

Query Match 87.6%; Score 18.4; DE 10; Length 153955;
Best Local Similarity 95.0%; Pred. No. 1.2e-02; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TGTGCGGGAGTTATTGAGT 21
Db 61061 TGTGCGGGAGTTATTGAGT 61080

RESULT 3
AL353093

LOCUS Human DNA sequence from clone RP11-224P11 on chromosome 10,
DEFINITION complete sequence.
AL353093
VERSION AL353093.18 GI:16943977
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Lovell, J.

Direct Submission
Submitted (07-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk

On Nov 15, 2001 this sequence version replaced gi:15131975.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30);
an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr10> RP11-224P11 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6
IMPORTANT: This sequence is not the entire insert of clone RP11-224P11 it may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true right end of clone RP11-224P11 is at 156008 in this sequence. The true left end of clone RP11-9P19 is at 131587 in this sequence. The true right end of clone RP11-192N10 is at 2000 in this sequence.

FEATURES

Location/Qualifiers
1..156008
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-224P11"
/clone_lib="RPCI-11.1"

ORIGIN

Query Match 87.6%; Score 18.4; DB 9; Length 156008;
Best Local Similarity 95.0%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 TGTGGGGAGTATTGAGT 21
|||
Db 25133 TGGGTGGGGAGTATTGAGT 25152

RESULT 4

AC073258 156127 bp DNA linear PRI 10-MAY-2001
LOCUS AC073258
DEFINITION Homo sapiens BAC clone RP11-221B19 from 7, complete sequence.
ACCESSION AC073258
VERSION AC073258.9 GI:13162541
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 156127)
AUTHORS Bielicki, L., Abbott, A. and Tonn, M.
TITLE The sequence of Homo sapiens BAC clone RP11-221B19
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 156127)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE

AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (28-FEB-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE

5 (bases 1 to 156127)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
6 (bases 1 to 156127)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (10-MAY-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Feb 28, 2001 this sequence version replaced gi:11995644.

REFERENCE

AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (10-MAY-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Feb 28, 2001 this sequence version replaced gi:11995644.

COMMENT

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.edu
----- Summary Statistics

Center project name: H_NH0221B19

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GTB/CHR7>, send mailto:egreen@nhgri.nih.gov, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Taten, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-115D14; the clone sequenced to the right is GS1-120K9, 200 bp overlap. Actual start of this clone is at base position 1 of RP11-221B19; actual end is at base position 13333 of GS1-120K9.

FEATURES

source
1..156127
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="7"
/map="7"
/clone="RP11-221B19"
/clone_lib="RPCI-11"
77..172
/rpt_family="CT-rich"
277..1516
/rpt_family="L1"
1570..1734
/rpt_family="L1"
repeat_region
repeat_region
repeat_region


```

repeat_region 2385..2413 /rpt_family="AT_rich"
repeat_region 2322..2351 /rpt_family="AT_rich"
repeat_region 3147..3282 /rpt_family="L1"
repeat_region 3567..3589 /rpt_family="AT_rich"
repeat_region 3637..3658 /rpt_family="AT_rich"
repeat_region 3671..3703 /rpt_family="AT_rich"
repeat_region 4374..4408 /rpt_family="AT_rich"
repeat_region 5097..5167 /rpt_family="AT_rich"
repeat_region 5245..5697 /rpt_family="ERV1"
repeat_region 6444..6722 /rpt_family="L2"
repeat_region 7050..7656 /rpt_family="L2"
repeat_region 7658..8270 /rpt_family="ACHobo"
repeat_region 8439..8597 /rpt_family="ACHobo"
repeat_region 8653..8776 /rpt_family="ACHobo"
repeat_region 9433..9463 /rpt_family="Alu"
repeat_region 10271..10330 /rpt_family="AT_rich"
repeat_region 10871..11339 /rpt_family="AT_rich"
repeat_region 11535..11931 /rpt_family="L2"
repeat_region 11932..13649 /rpt_family="MALR"
repeat_region 13651..13754 /rpt_family="MALR"
repeat_region 13759..13953 /rpt_family="MALR"
repeat_region 13964..14359 /rpt_family="MALR"
repeat_region 16583..16859 /rpt_family="L1"
repeat_region 16903..16949 /rpt_family="(TA)n"
repeat_region 16950..16973 /rpt_family="(TG)n"
repeat_region 17371..17549 /rpt_family="MIR"
repeat_region 18153..18173 /rpt_family="AT_rich"
repeat_region 18460..18598 /rpt_family="T2_type"
repeat_region 18719..18743 /rpt_family="(TG)n"
repeat_region 19092..19126 /rpt_family="AT_rich"
repeat_region 19164..19212 /rpt_family="AT_rich"
repeat_region 19740..22078 /rpt_family="L1"
repeat_region 23259..23289 /rpt_family="AT_rich"
repeat_region 23616..23653 /rpt_family="AT_rich"
repeat_region 23716..24002 /rpt_family="Alu"
misc_feature 24118..24219 /note="similar to EST AV707173 (NID:gi0724443)"
misc_feature 24274..24846

```

```

/note="similar to EST AV707173 (NID:gi0724443)"
25095..25160 /rpt_family="(TG)n"
25669..25844 /rpt_family="L2"
27560..27601 /rpt_family="AT_rich"
27649..27748 /rpt_family="L1"
27749..27801 /rpt_family="(TA)n"
27802..27837 /rpt_family="L1"
28596..28646 /rpt_family="MIR"
28606..29519 /rpt_family="L2"
29677..29715 /rpt_family="(CA)n"
29790..29905 /rpt_family="L2"
31210..31407 /rpt_family="MIR"

```

Query Match 87.6%; Score 18.4; DB 9; Length 156127;
 Best Local Similarity 95.0%; Pred. No. 1.2e+02;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Cy 1 TTGTGGGGAGTTATTGAG 20
 |||||
 Db 142898 TTGTGGGGAGTTATTGAG 142917

RESULT 5

AC013555/c

LOCUS

DEFINITION AC013555 183251 bp DNA linear HTG 09-MAY-2001
 Homo sapiens chromosome 18 clone RP11-2011 map 18, WORKING DRAFT
 SEQUENCE, 30 unordered pieces.

AC013555

VERSION AC013555.4 GI:9369454

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE
 1 (Bases 1 to 183251)
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 Homo sapiens chromosome 18, clone RP11-2011

TITLE
 JOURNAL
 AUTHORS

REFERENCE
 2 (Bases 1 to 183251)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
 Baldwin, J., Barna, N., Beckery, R., Boguslavsky, L., Bouckgatter, B.,
 Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,
 Cooke, P., Dearellano, K., Dewar, K., Domino, M., Donegan, L., Doyle, M.,
 Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,
 Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
 Lenoczky, J., Lien, C., Locke, K., MacDonald, P., Marquis, N.,
 McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meidrim, J.,
 Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
 Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
 Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (13-NOV-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jul 22, 2000 this sequence version replaced gi:8567844.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE

JOURNAL

COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: W18R
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information -----
Center project name: I2231
Center clone name: 2.O.11
----- Summary Statistics -----
Sequencing vector: M13; W7815; 99% of reads
Sequencing vector: Plasmid; n/a; 0.0% of reads
1.3502779841144Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 163759 bases at least Q40
Consensus quality: 172974 bases at least Q30
Consensus quality: 177094 bases at least Q20
Insert size: 94000; agarose-fp
Insert size: 180351; sum-of-ctngs
Quality coverage: 5.6 in Q20 bases; agarose-fp
Quality cover.

NOTE: This is a 'working draft' sequence. It currently consists of 30 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

	1	1081	contig of 1081 bp in length
*	1082	1181	gap of 100 bp
*	1182	2405	contig of 1224 bp in length
*	2406	2505	gap of 100 bp
*	2506	3901	contig of 1396 bp in length
*	3902	4001	gap of 100 bp
*	4002	5286	contig of 1285 bp in length
*	5287	5386	gap of 100 bp
*	5387	7468	contig of 2082 bp in length
*	7469	7568	gap of 100 bp
*	7569	9376	contig of 1808 bp in length
*	9377	9476	gap of 100 bp
*	9477	10817	contig of 1341 bp in length
*	10818	10917	gap of 100 bp
*	10918	13577	contig of 2650 bp in length
*	13578	13677	gap of 100 bp
*	13678	16194	contig of 2517 bp in length
*	16195	16294	gap of 100 bp
*	16295	18903	contig of 2609 bp in length
*	18904	19003	gap of 100 bp
*	19004	21391	contig of 2388 bp in length
*	21392	21491	gap of 100 bp
*	21492	25590	contig of 4099 bp in length
*	25591	25830	gap of 100 bp
*	25831	28465	contig of 2775 bp in length
*	28466	28565	gap of 100 bp
*	28566	31371	contig of 2806 bp in length
*	31372	31471	gap of 100 bp
*	31472	34943	contig of 3472 bp in length
*	34944	35043	gap of 100 bp
*	35044	38235	contig of 3192 bp in length
*	38236	38335	gap of 100 bp
*	38336	40221	contig of 2586 bp in length
*	40222	41021	gap of 100 bp
*	41022	45666	contig of 4645 bp in length
*	45667	45766	gap of 100 bp
*	45767	50004	contig of 4238 bp in length
*	50005	50104	gap of 100 bp
*	50105	56058	contig of 5954 bp in length
*	56059	56158	gap of 100 bp
*	56159	62036	contig of 5878 bp in length
*	62037	62136	gap of 100 bp
*	62137	68894	contig of 6758 bp in length
*	68895	68994	gap of 100 bp
*	68995	75499	contig of 6504 bp in length
*	75499	75598	gap of 100 bp

	75599	82295	contig of 6697 bp in length
*	82296	82395	gap of 100 bp
*	82396	90453	contig of 8058 bp in length
*	90454	90553	gap of 100 bp
*	90554	101085	contig of 10532 bp in length
*	101086	101186	gap of 100 bp
*	101186	111226	contig of 10041 bp in length
*	111227	111227	gap of 100 bp
*	111327	122293	contig of 10967 bp in length
*	122294	122393	gap of 100 bp
*	122394	149176	contig of 26783 bp in length
*	149177	149276	gap of 100 bp
*	149277	183251	contig of 33975 bp in length

FEATURES
Location/Qualifiers
source
1. 183251
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-2011"
/clone_lib="RPC1-11 Human Male BAC"
1. 1081
/note="assembly_fragment"
1182. 2405
/note="assembly_fragment"
2506. 3901
/note="assembly_fragment"
4002. 5286
/note="assembly_fragment"
5387. 7468
/note="assembly_fragment"
7569. 9376
/note="assembly_fragment"
9477. 10817
/note="assembly_fragment"
vector_side:right
10918. 13577
/note="assembly_fragment"
13678. 16194
/note="assembly_fragment"
16295. 18903
/note="assembly_fragment"
19004. 21391
/note="assembly_fragment"
21492. 25590
/note="assembly_fragment"
25691. 28465
/note="assembly_fragment"
28566. 31371
/note="assembly_fragment"
31472. 34943
/note="assembly_fragment"
35044. 38235
/note="assembly_fragment"
38336. 40921
/note="assembly_fragment"
41022. 45666
/note="assembly_fragment"
45767. 50004
/note="assembly_fragment"
clone_end:SP6
vector_side:right
50105. 56058
/note="assembly_fragment"
56159. 62036
/note="assembly_fragment"
62137. 68894
/note="assembly_fragment"
68995. 75498
/note="assembly_fragment"
75599. 82295

```

/note="assembly_fragment"
82396. .90453
/note="assembly_fragment"
90554. .101085
/note="assembly_fragment"
101186. .111226
/note="assembly_fragment"
111327. .122293
/note="assembly_fragment"
122394. .149176
/note="assembly_fragment"
149277. .183251
/note="assembly_fragment"

ORIGIN
Query Match      87.6%; Score 18.4; DB 2; Length 183251;
Best Local Similarity 95.0%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTGTGTGGGGAGTTATTGAG 20
    |||||
Db 156770 TTGTGTGGGGGTATTGAG 156751
    |||||

RESULT 6
AC025105/c
LOCUS      201027 bp      DNA      linear      PRI 20-JUL-2002
DEFINITION Homo sapiens chromosome 18, clone RP11-695P1, complete sequence.
ACCESSION  AC025105
VERSION    AC025105.10 GI:21913036
KEYWORDS   HTG.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 201027)
AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
TITLE     Homo sapiens chromosome 18, clone RP11-695P1
JOURNAL   Unpublished
REFERENCE  2 (bases 1 to 201027)
AUTHORS   Birren,B., Linton,J., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,C.,
Pisani,C., Pollara,V., Raymond,C., Riley,K., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (05-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 201027)
REFERENCE  3 (bases 1 to 201027)
AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Canarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,

```

```

Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (27-JUN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 201027)
REFERENCE  4 (bases 1 to 201027)
AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Canarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (20-JUL-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 20, 2002 this sequence version replaced gi:21617749.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WITB
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5654
Center clone name: 695_P1
----- Location/Qualifiers
1. 201027
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-695P1"
/clone_lib="RPC1-11 Human Male BAC"
repeat_region 46. .180
/rpt_family="L1"
repeat_region 245. .433
/rpt_family="L1"
repeat_region 480. .521
/rpt_family="(CA)n"
repeat_region 530. .680
/rpt_family="L1MC/D"
repeat_region 1191. .1477
/rpt_family="AluY"
repeat_region complement(1783. .1981)
/rpt_family="WERS58A"
repeat_region complement(2294. .2646)
/rpt_family="AluSp"
repeat_region 3197. .3315
/rpt_family="L3"

```

```

repeat_region 3367..3448
/rpt_family="MIR"
complement(3796..3966)
repeat_region /rpt_family="MER117"
4157..4272
/rpt_family="TTTTn"
complement(4274..4541)
repeat_region /rpt_family="AluSx"
complement(4787..5229)
repeat_region /rpt_family="MLT1C"
complement(5317..5321)
unsure /note="<30 qual SNGL region"
5480..5839
repeat_region /rpt_family="MER77"
6137..6208
repeat_region /rpt_family="CT-rich"
complement(6904..8908)
unsure /note="<30 qual SNGL region"
complement(9002..9006)
unsure /note="<30 qual SNGL region"
complement(9050..9051)
unsure /note="probably TC; possibly TAC"
complement(9313..9319)
unsure /note="<30 qual SNGL region"
9533..9638
repeat_region /rpt_family="(TG)n"
10850..10856
unsure /note="<30 qual SNGL region"
10916..10926
unsure /note="<30 qual SNGL region"
10918..10919
unsure /note="probably CC, possibly CCC"
10965..10970
unsure /note="<30 qual SNGL region"
11016..11034
unsure /note="<30 qual SNGL region"
11066..11070
unsure /note="<30 qual SNGL region"
11190..11195
unsure /note="<30 qual SNGL region"
12241..12405
repeat_region /rpt_family="(TG)n"
12440..12500
repeat_region /rpt_family="(TG)n"
12519..12694
repeat_region /rpt_family="(TG)n"
12712..12794
repeat_region /rpt_family="(TATATG)n"
12801..12951
repeat_region /rpt_family="(TG)n"
13673..13711
repeat_region /rpt_family="(T)n"
16731..16821
repeat_region /rpt_family="(TG)n"
16862..17041
repeat_region /rpt_family="(TG)n"
17971..18181
repeat_region /rpt_family="MLT2C2"
complement(18182..18196)
repeat_region /rpt_family="FLAM_C"
complement(18197..18320)
repeat_region /rpt_family="LIP4"
complement(18321..18431)
repeat_region /rpt_family="FLAM_C"
18459..18570
repeat_region /rpt_family="MLT2C2"
18571..18874
repeat_region /rpt_family="AluSx"
18876..18895
repeat_region /rpt_family="(TC)n"
18898..18931
repeat_region /rpt_family="(CA)n"
18932..18960
/rpt_family="MLT2C2"
complement(20168..20228)
/rpt_family="MLT1A1"
20229..20539
/rpt_family="AluSx"
complement(20540..20859)
/rpt_family="MLT1A1"

Query Match 87.6%; Score 18.4; DB 9; Length 201027;
Best Local Similarity 95.0%; Pred.No.1.le-02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTGTGTGGGGAGTTATTGAG 20
|||||
Db 85987 TTGTGTGGGGTGTATTGAG 85968

RESULT 7
AL591706 201725 bp DNA linear HTG 19-DEC-2001
LOCUS Homo sapiens chromosome 10 clone RP13-426G13, WORKING DRAFT
DEFINITION SEQUENCE, 10 unordered pieces.
ACCESSION AL591706
VERSION AL591706.1 GI:14270168
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1 Plumb, B.
AUTHORS Direct Submission
TITLE Submitted (18-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,
JOURNAL Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT ----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BB426G13
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 199206 bases at least Q40
Consensus quality: 199917 bases at least Q30
Consensus quality: 200376 bases at least Q20
Insert size: 200825; sum-of-contigs
Quality coverage: 7.92x in Q20 bases; sum-of-contigs Quality
Coverage: 8.87x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 11133: contig of 11133 bp in length
* 11134 11233: gap of 100 bp
* 11234 44021: contig of 32788 bp in length
* 44022 44121: gap of 100 bp
* 44122 74520: contig of 30399 bp in length
* 74521 74621: gap of 100 bp
* 74621 108523: contig of 33902 bp in length
* 108523 120300: gap of 100 bp
* 108623 120300: contig of 11578 bp in length
* 120201 120300: gap of 100 bp
* 120301 123065: contig of 2765 bp in length

```

```
* 123066 123165: gap of 100 bp
* 123166 132118: contig of 9153 bp in length
* 132118 132119: gap of 100 bp
* 132119 168510: contig of 36092 bp in length
* 168510 168611: gap of 100 bp
* 168611 187993: contig of 19383 bp in length
* 187993 188093: gap of 100 bp
* 188093 201725: contig of 13632 bp in length.
```

FEATURES

```
Location/Qualifiers
1..201725
  /organism="Homo sapiens"
  /mol_type="genomic DNA"
  /db_xref="taxon:9606"
  /chromosome="10"
  /clone="RP13-426G13"
  /clone_lib="RPC1-13.2"
1..11133
  /note="assembly_fragment:03649"
  /vector="pGEM-1"
  /vector_size:left
11234..44021
  /note="assembly_fragment:03972"
44122..74520
  /note="assembly_fragment:04102"
  /fragment_chain:1
74621..108522
  /note="assembly_fragment:01076"
  /fragment_chain:1
108623..120200
  /note="assembly_fragment:03714"
  /fragment_chain:1
120301..123065
  /note="assembly_fragment:03751"
  /fragment_chain:1
123166..132318
  /note="assembly_fragment:02246"
  /fragment_chain:1
132419..168510
  /note="assembly_fragment:03426"
  /fragment_chain:1
168611..187993
  /note="assembly_fragment:03239"
  /fragment_chain:1
188094..201725
  /note="assembly_fragment:02682"
  /fragment_chain:1
  /clone_end:T7
  /vector_size:right
```

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

ORIGIN

Query Match 87.6%; Score 18.4; DB 2; Length 201725;

Best Local Similarity 95.0%; Pred. No. 1.1e+02;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TGTGTGGGGAGTTATTCAGT 21

|||||

Db 57830 TGGGTGGGGAGTTATTCAGT 57849

RESULT 8

```
AC107502 AC107502 223768 bp DNA linear HTG 13-MAY-2003
LOCUS Rattus norvegicus clone CH230-127P6, WORKING DRAFT SEQUENCE.
DEFINITION AC107502
ACCESSION AC107502.5 GI:30580542
VERSION HTG; HTGS PHASE2; HTGS DRAFT; HTGS_FULLTOP.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
```

ORGANISM

```
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 223768)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Aider,J.,
```

Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Ayoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escoto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerrin,I., Guevara,W., Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhera,L., Loulseghe,H., Lozano,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Hawthiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,S., Munitasa,M., Murphy,M., Nair,L., Morgan,M., Morris,K., Morris,S., Munitasa,M., Nguyen,N., Norris,S., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K., Nwankwelen,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L., Pu,L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,P., Rives,C., Rodkey,I., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajds,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasaana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wlaczek,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

Direct Submission

Unpublished

2 (bases 1 to 223768)

Worley,K.C.

Direct Submission

Submitted (23-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 223768)

Rat Genome Sequencing Consortium

Direct Submission

Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On May 13, 2003 this sequence version replaced gi:22855792.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be whole contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome

shotgun sequence only contigs will be indicated in the feature table.

```

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
-----
Center project name: GJAV
Center clone name: CH230-127F6
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 212851 bases at least Q40
Consensus quality: 215410 bases at least Q30
Consensus quality: 216880 bases at least Q20
Estimated insert size: 221458; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
  consists of 1 contigs. Gaps between the contigs
  are represented as runs of N. The order of the pieces
  is believed to be correct as given, however the sizes
  of the gaps between them are based on estimates that have
  been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
  the accession number will be preserved.
* 1 223768: contig of 223768 bp in length.
  Location/Qualifiers
    1..223768
     /organism="Rattus norvegicus"
     /mol_type="genomic DNA"
     /db_xref="taxon:10116"
     /clone="CH230-127F6"
     216537..218381
      /note="wgs contig"
     218861..220305
      /note="wgs contig"
     222567..223768
      /note="wgs contig"
ORIGIN
Query Match      87.6%; Score 18.4; DB 2; Length 223768;
Best Local Similarity 95.0%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TGTGTGGGGAGTTAGT 21
Db 114893 TGTGTGGGGAGTTAGT 114912

RESULT 9
AC115140/c
LOCUS
DEFINITION
Rattus norvegicus clone CH230-99M12, *** SEQUENCING IN PROGRESS
***, 4 unordered pieces.
ACCESSION
AC115140
VERSION
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS
Rattus norvegicus (Norway rat)
SOURCE
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 260335)
Muzny, D., Marie, J., Metzker, M., Lee, J., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Albrechts, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Ayodeji, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM

```

```

Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, J., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregiorgis, B., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogue, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idels, D., Jackson, A.,
Jacks, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowals, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenshew, L., Louisa, H., Lozada, R. J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munitasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwankweme, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.,
Puro, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smales, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usman, K.,
Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
Weinstock, G. and Gibbs, R. A.
Direct Submission
Unpublished
2 (bases 1 to 260335)
Worley, K. C.
Direct Submission
Submitted (15-MAR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 260335)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (08-OCT-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 23, 2002 this sequence version replaced gi:21736944.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM

```

Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GNXU
 Center clone name: CH230-99N12

----- Summary Statistics

Assembly program: Phrap; version 0.990329
 Consensus quality: 222523 bases at least Q40
 Consensus quality: 225495 bases at least Q30
 Consensus quality: 227152 bases at least Q20
 Estimated insert size: 239709; sum-of-contigs estimation
 Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 (see <http://www.hgsc.bcm.tmc.edu/docs/Genbankdraftdata.html>).
 * NOTE: This is a 'working draft' sequence. It currently
 consists of 4 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

* 1 172711: contig of 172711 bp in length
 * 172712 172811: gap of unknown length
 * 172812 191948: contig of 19137 bp in length
 * 191949 192048: gap of unknown length
 * 192049 258860: contig of 68812 bp in length
 * 258861 258960: gap of unknown length
 * 258961 260335: contig of 1375 bp in length.

FEATURES

source
 1..260335
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-99N12"

misc_feature
 1..1294

misc_feature
 175748..176884

/note="wgs_contig"
 /note="wgs_contig"

ORIGIN

Query Match 85.7%; Score 18; DB 2; Length 260335;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 TTGTGGGGAGTTATTGAGT 21

Db 221219 TTGTGGGGAGTTATTGAGT 221202

RESULT 10

AX822420 12781 bp DNA linear PAT 11-DEC-2003
 LOCUS
 DEFINITION Sequence 312 from Patent EPI340818.
 ACCESSION AX822420
 VERSION AX822420.1 GI:39749048
 KEYWORDS
 SOURCE synthetic construct
 ORGANISM synthetic construct
 artificial sequences.

REFERENCE

1
 AUTHORS Adorjan,P., Burger,M., Maier,S., Nimmrich,I., Becker,E., Lesche,R.,
 Rujan,T. and Schmitt,A.
 TITLE Method and nucleic acids for the analysis of a colon cell
 proliferative disorder
 JOURNAL Patent: EP 1340818-A 312 03-SEP-2003;
 Epigenomics AG (DE)

FEATURES

source
 1..12781
 /organism="synthetic construct"
 /mol_type="unassigned DNA"
 /db_xref="taxon:132630"

ORIGIN
 /note="chemically treated genomic DNA (Homo sapiens)"

Query Match 84.8%; Score 17.8; DB 6; Length 12781;
 Best Local Similarity 90.5%; Pred. No. 5.1e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTGTGTGGGGAGTTATTGAGT 21

Db 7985 TTGTGTGGGGAGTTATTGAGT 8005

RESULT 11

AX826060 12781 bp DNA linear PAT 11-DEC-2003
 LOCUS
 DEFINITION Sequence 312 from Patent WO03072821.
 ACCESSION AX826060
 VERSION AX826060.1 GI:39751574
 KEYWORDS
 SOURCE synthetic construct
 ORGANISM synthetic construct
 artificial sequences.

REFERENCE

1
 AUTHORS Adorjan,P., Burger,M., Maier,S., Nimmrich,I., Becker,E., Lesche,R.,
 Rujan,T. and Schmitt,A.
 TITLE Method and nucleic acids for the analysis of a colon cell
 proliferative disorder
 JOURNAL Patent: WO 03072821-A 312 04-SEP-2003;
 Epigenomics AG (DE)

FEATURES

source
 1..12781
 /organism="synthetic construct"
 /mol_type="unassigned DNA"
 /db_xref="taxon:32630"
 /note="chemically treated genomic DNA (Homo sapiens)"

ORIGIN

Query Match 84.8%; Score 17.8; DB 6; Length 12781;
 Best Local Similarity 90.5%; Pred. No. 5.1e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTGTGTGGGGAGTTATTGAGT 21

Db 7985 TTGTGTGGGGAGTTATTGAGT 8005

RESULT 12

AX346237 19653 bp DNA linear PAT 01-FEB-2002
 LOCUS
 DEFINITION Sequence 1308 from Patent WO0200928.
 ACCESSION AX346237
 VERSION AX346237.1 GI:18494123
 KEYWORDS
 SOURCE synthetic construct
 ORGANISM synthetic construct
 artificial sequences.

REFERENCE

1
 AUTHORS Olek,A., Piepenbrock,C. and Berlin,K.
 TITLE Diagnosis of diseases associated with the immune system
 JOURNAL Patent: WO 0200928-A 1308 03-JAN-2002;
 Epigenomics AG (DE)

FEATURES

source
 1..19653
 /organism="synthetic construct"
 /mol_type="unassigned DNA"
 /db_xref="taxon:32630"
 /note="chemically treated genomic DNA (Homo sapiens)"

ORIGIN

Query Match 84.8%; Score 17.8; DB 6; Length 19653;
 Best Local Similarity 90.5%; Pred. No. 4.4e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY      1  TTGTGTGGGAGTTATTGAGT 21
Db      17155 TGGTGTGGGAGTTATAGAGT 17175

RESULT 13
AL732359.00/c
WFCOMMENT
Sequence split into 16 fragments LOCUS AL732359 Accession AL732359
Fragment Name      Begin      End
AL732359_00        1        110000
AL732359_01        100001    210000
AL732359_02        200001    310000
AL732359_03        300001    410000
AL732359_04        400001    510000
AL732359_05        500001    610000
AL732359_06        600001    710000
AL732359_07        700001    810000
AL732359_08        800001    910000
AL732359_09        900001   1010000
AL732359_10       1000001   1100000
AL732359_11       1100001   1210000
AL732359_12       1200001   1310000
AL732359_13       1300001   1410000
AL732359_14       1400001   1510000
AL732359_15       1500001   1571086
LOCUS   AL732359      1571086 bp    DNA      linear      HTG 04-NOV-2003
DEFINITION Homo sapiens chromosome 6 clone XXYac-82D3, WORKING DRAFT SEQUENCE,
31 unordered pieces.
ACCESSION AL732359
VERSION   AL732359.23 GI:38175013
KEYWORDS  HTG; HTGS PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1571086)
AUTHORS   Wood, J.
TITLE     Direct Submission
JOURNAL   Submitted (03-NOV-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgehire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Nov 4, 2003 this sequence version replaced gi:35209238.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: YX82D3
----- Summary Statistics
Sequencing vector: XGAP4; version 4.5
Chemistry: Dye-terminator; 97% of reads
Chemistry: Dye-terminator Big Dye; 2% of reads
Consensus quality: 1291964 bases at least Q40
Consensus quality: 1295639 bases at least Q30
Consensus quality: 1297808 bases at least Q20
Insert size: 1568086; sum-of-contigs
Quality coverage: 6.71x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 31 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 12202: contig of 12202 bp in length
* 12203 12302: gap of 100 bp
* 12303 17846: contig of 5544 bp in length
* 17847 17946: gap of 100 bp

```

```

17947 50950: contig of 33004 bp in length
50951 51050: gap of 100 bp
51051 58728: contig of 7678 bp in length
58729 58828: gap of 100 bp
58829 67002: contig of 8174 bp in length
67003 67102: gap of 100 bp
67103 92125: contig of 25023 bp in length
92126 92225: gap of 100 bp
92226 95249: contig of 3024 bp in length
95250 95349: gap of 100 bp
95350 245613: contig of 150264 bp in length
245614 245713: gap of 100 bp
245714 472564: contig of 228851 bp in length
472565 472665: gap of 100 bp
472666 505271: contig of 32607 bp in length
505272 505371: gap of 100 bp
505372 559591: contig of 54220 bp in length
559592 559691: gap of 100 bp
559692 873847: contig of 314156 bp in length
873848 873947: gap of 100 bp
873948 986519: contig of 112572 bp in length
986520 986620: gap of 100 bp
986621 1180781: contig of 194162 bp in length
1180782 1218552: contig of 37671 bp in length
1218553 1218552: gap of 100 bp
1218553 1228314: contig of 9662 bp in length
1228315 1228414: gap of 100 bp
1228415 1278395: contig of 49981 bp in length
1278396 1278495: gap of 100 bp
1278496 1281408: contig of 2913 bp in length
1281409 1281508: gap of 100 bp
1281509 1283777: contig of 2269 bp in length
1283778 1283877: gap of 100 bp
1283878 1286039: contig of 2162 bp in length
1286040 1286139: gap of 100 bp
1286140 1288220: contig of 2081 bp in length
1288221 1288320: gap of 100 bp
1288321 1290462: contig of 2142 bp in length
1290463 1290562: gap of 100 bp
1290563 1292742: contig of 2180 bp in length
1292743 1292842: gap of 100 bp
1292843 1294508: contig of 2086 bp in length
1294509 1295008: gap of 100 bp
1295009 1297355: contig of 2347 bp in length
1297356 1297455: gap of 100 bp
1297456 1299648: contig of 2193 bp in length
1299649 1299748: gap of 100 bp
1299749 1302495: contig of 2747 bp in length
1302496 1302595: gap of 100 bp
1302596 1304639: contig of 2044 bp in length
1304640 1304739: gap of 100 bp
1304740 1306861: contig of 2122 bp in length
1306862 1306961: gap of 100 bp
1306962 1413469: contig of 106508 bp in length
1413470 1413569: gap of 100 bp
1413570 1571086: contig of 157517 bp in length.
Location/Qualifiers
1. .1571086
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="6"
/clone="XXYac-82D3"
/clone_lib="Human-ICRF-YAC"
1. .12202
/note="assembly fragment:19199
fragment chain:1"
12303. 17846
/note="assembly fragment:02286
fragment chain:1"
17947. 50950
/note="assembly_fragment:13330"

```

```

FEATURES
Source
1. .1571086
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="6"
/clone="XXYac-82D3"
/clone_lib="Human-ICRF-YAC"
1. .12202
/note="assembly fragment:19199
fragment chain:1"
12303. 17846
/note="assembly fragment:02286
fragment chain:1"
17947. 50950
/note="assembly_fragment:13330"

```



```

misc_feature      fragment chain:1"
51051. 58728
/note="assembly fragment:17055
fragment chain:1"
58829. 67002
/note="assembly fragment:07211
fragment chain:1"
67103. 92125
/note="assembly fragment:00427
fragment chain:1"
92226. 95249
/note="assembly fragment:28324
fragment chain:1"
95350. 245613
/note="assembly fragment:13688
fragment chain:2"
245714. 472564
/note="assembly fragment:12100
fragment chain:2"
47265. 505271
/note="assembly fragment:27850
fragment chain:2"
505372. 559591
/note="assembly fragment:03772
fragment chain:2"
559692. 873847
/note="assembly fragment:29257
fragment chain:3"
873948. 986519
/note="assembly fragment:04215
fragment chain:3"
986620. 1180781
/note="assembly fragment:24107
fragment chain:3"
1180852. 1218552
/note="assembly fragment:23115
fragment chain:3"
1218653. 1228314
/note="assembly fragment:06423
fragment chain:4"
1228415. 1278395
/note="assembly fragment:01809
fragment chain:4"
1278496. 1281408
/note="assembly fragment:00009"
1281509. 1283777
/note="assembly fragment:08095"
1283878. 1286039
/note="assembly fragment:11475"
1286140. 1288220
/note="assembly fragment:13091"
1288321. 1290462
/note="assembly fragment:14466"
1290563. 1292742
/note="assembly fragment:19168"
1292843. 1294908
/note="assembly fragment:19437"
1295009. 1297355
/note="assembly fragment:21640"
1297456. 1299648
/note="assembly fragment:22679"

```

Query Match 84.8%; Score 17.8; DB 2; Length 110000;
 Best Local Similarity 90.5%; Pred. No. 2.6e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 TTGTGCGGGAGTTATTGAGT 21
    |||||
DB 45412 TTGTGCGGGAGTTATTGAGT 45392
    |||||

```

RESULT 14
 BX005055

```

LOCUS              BX005055              156598 bp      DNA      linear      VFT 23-SEP-2003
DEFINITION          Zebrafish DNA sequence from clone CH211-188N23 in linkage group 23,
                     complete sequence.
ACCESSION            BX005055
VERSION              BX005055.5  GI:34996458
KEYWORDS              HTG
SOURCE                Danio rerio (zebrafish)
ORGANISM              Danio rerio
                     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                     Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
                     Cypriniformes; Cyprinidae; Danio.
REFERENCE             1. (bases 1 to 156598)
AUTHORS               Whitehead, S.
TITLE                 Direct Submission
JOURNAL               Submitted (23-SEP-2003) Wellcome Trust Sanger Institute, Hinxton,
                     Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
                     zfsh-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT               On Sep 23, 2003 this sequence version replaced gi:33412026.
                     ----- Genome Center
                     Center: Wellcome Trust Sanger Institute
                     Web site: http://www.sanger.ac.uk
                     Contact: zfsh-help@sanger.ac.uk
                     -----
                     During sequence assembly data is compared from overlapping clones.
                     Where differences are found these are annotated as variations
                     together with a note of the overlapping clone name. Note that the
                     variation annotation may not be found in the sequence submission
                     corresponding to the overlapping clone, as we submit sequences with
                     only a small overlap as described above.
                     This sequence was finished as follows unless otherwise noted: all
                     regions were either double-stranded or sequenced with an alternate
                     chemistry or covered by high quality data (i.e., phred quality >=
                     30); an attempt was made to resolve all sequencing problems, such
                     as compressions and repeats; all regions were covered by at least
                     one plasmid subclone or more than one M13 subclone; and the
                     assembly was confirmed by restriction digest, except on the rare
                     occasion of the clone being a YAC.
                     The following abbreviations are used to associate primary accession
                     numbers given in the feature table with their source databases:
                     Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
                     on the WORMPEP database can be found at
                     http://www.sanger.ac.uk/projects/C.elegans/wormpep/Clone-derived
                     Zebrafish pUC subclones occasionally display inconsistency over the
                     length of mononucleotide A/T runs and conserved TA repeats. Where
                     this is found the longest good quality representation will be
                     submitted.
                     Repeat names beginning 'Dr' were identified by the Recon repeat
                     discovery system (Zhihong Bao and Sean Eddy, submitted), and those
                     beginning 'drr' were identified by Rick Waterman (Stephen Johnson
                     lab, WashU) For further information see
                     http://www.sanger.ac.uk/projects/D.rerio/fishmask.shtml
                     CH211-188N23 is from a CHORI-211 BAC library
                     VECTOR: PTARBAC2.1.
                     Location/Qualifiers
                     1. 156598
                        /organism="Danio rerio"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:7955"
                        /clone="CH211-188N23"
                        /clone_lib="CHORI-211"

```

FEATURES source

ORIGIN

Query Match 84.8%; Score 17.8; DB 5; Length 156598;
 Best Local Similarity 90.5%; Pred. No. 2.3e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 TTGTGCGGGAGTTATTGAGT 21
    |||||
DB 150294 TTGTGCGGGAGTTATTGAGT 150314
    |||||

```

RESULT 15

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2004, 17:01:15 ; Search time 177.593 Seconds
(without alignments)
502.342 Million cell updates/sec

Title: US-10-084-555A-115

Perfect score: 21

Sequence: 1 ttgtgtgggagtgattgagt 21

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_29Jan04:*

1: Geneseq1980s:*

2: Geneseq1990s:*

3: Geneseq2000s:*

4: Geneseq2001as:*

5: Geneseq2001bs:*

6: Geneseq2002s:*

7: Geneseq2003as:*

8: Geneseq2003bs:*

9: Geneseq2003cs:*

10: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17.8	84.8	12781	9	ADB54256
2	17.8	84.8	19653	6	ABL33335
3	16.8	80.0	730	5	AAS73256
4	16.8	80.0	2711	5	AAS79261
5	16.8	80.0	2764	2	AAT43351
6	16.8	80.0	5557	6	ABL33547
7	16.8	80.0	5756	6	ABL32586
8	16.8	80.0	50000	9	ADC87687
9	16.8	80.0	75899	6	ABK85261
10	16.4	78.1	1932	9	ADE56603
11	16.4	78.1	1932	9	ADE56607
12	16.4	78.1	1932	9	ADE56611
13	16.4	78.1	1932	9	ADE56599
14	16.4	78.1	5259	6	ABQ68985
15	16.4	78.1	8410	4	ABL22880
16	16.2	77.1	587	6	ABQ35536
17	16.2	77.1	1200	6	ABQ35537
18	16.2	77.1	1200	6	ABQ18222
19	16.2	77.1	1200	6	ABQ18223
20	16.2	77.1	1201	6	ABQ14377
21	16.2	77.1	1201	6	ABQ14376
22	16.2	77.1	1413	6	ABQ32828
23	16.2	77.1	1413	6	ABQ32829

c	24	16.2	77.1	1418	6	ABQ32843
c	25	16.2	77.1	1418	6	ABQ32842
c	26	16.2	77.1	1650	4	ABL21131
c	27	16.2	77.1	2960	9	ADE55515
c	28	16.2	77.1	3121	4	ABL28812
c	29	16.2	77.1	4675	4	ABL28858
c	30	16.2	77.1	4937	2	AAQ49507
c	31	16.2	77.1	4937	2	AAQ72294
c	32	16.2	77.1	4937	2	AAQ72294
c	33	16.2	77.1	4940	3	AAQ52197
c	34	16.2	77.1	5070	6	ABL32801
c	35	16.2	77.1	5116	2	AAQ49506
c	36	16.2	77.1	5116	2	AAQ72293
c	37	16.2	77.1	5116	2	AAQ90994
c	38	16.2	77.1	5393	6	AAQ52195
c	39	16.2	77.1	5393	6	ABT07728
c	40	16.2	77.1	5488	6	ABL33456
c	41	16.2	77.1	5507	7	ABZ10204
c	42	16.2	77.1	5814	6	ABL33561
c	43	16.2	77.1	6063	6	ABK28394
c	44	16.2	77.1	6503	6	ABL32720
c	45	16.2	77.1	6618	6	ABK31284

ALIGNMENTS

RESULT 1

ADB54256

ID ADB54256 standard; DNA; 12781 BP.

XX

AC ADB54256;

XX

DT 04-DEC-2003 (first entry)

XX

DE Pretreated genomic DNA region 180.

XX

KW colon cell proliferative disorder; non methylated CpG dinucleotide;

KW cytostatic; cancer; adenoma; carcinoma; cytosine methylation state; ds.

XX

OS Unidentified.

XX

PN WO2003072821-A2.

XX

PD 04-SEP-2003.

XX

PF 27-FEB-2003; 2003WO-EP002035.

XX

PR 27-FEB-2002; 2002EP-00004551.

XX

PA (EPIG-) EPIGENOMICS AG.

XX

PI Adorjan P, Burger M, Maier S, Nimrich I, Becker E, Lesche R;

PI Rujan T, Schmitt A;

XX

DR WPI; 2003-731620/69.

XX

PT Detecting and differentiating between colon cell proliferative disorders associated with a gene or its regulatory regions comprises contacting a target nucleic acid in a biological sample obtained from the subject with a reagent.

XX

PS Claim 32; SEQ ID NO 312; 74pp; English.

XX

CC The invention relates to a novel method for detecting and differentiating between colon cell proliferative disorders associated with at least one gene or its regulatory regions. The method comprises contacting a target nucleic acid in a biological sample obtained from the subject with at least one reagent or a series of reagents, where the reagent or series of reagents, distinguishes between methylated and non methylated CpG dinucleotides within the target nucleic acid. The molecules of the invention demonstrate cytosine methylation activity whilst the method may be useful for detecting and differentiating between colon cell proliferative

CC

CC

CC

CC

CC

CC

CC

CC

CC disorders, including cancers such as colon adenoma and colon carcinoma.
 CC The PNA (peptide nucleic acid)-oligonucleotides are useful as probes for
 CC determining cytosine methylation state or single nucleotide
 CC polymorphisms. The current sequence is that of the pretreated genomic DNA
 CC region of the invention. This sequence is not shown within the
 CC specification but is taken from Wipoweb.

XX SQ Sequence 12781 BP; 3584 A; 0 C; 2893 G; 6304 T; 0 U; 0 Other;

Query Match 84.8%; Score 17.8; DB 9; Length 12781;

Best Local Similarity 90.5%; Pred. No. 81;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTGTGTGGGAGTTATTGAGT 21

DB 7985 TTGTGTGGGAGTTATTGAGT 8005

RESULT 2

ABL33335

ID ABL33335 standard; DNA; 19653 BP.

XX AC ABL33335;

XX DT 26-MAR-2002 (first entry)

XX DE Human immune system associated gene SEQ ID NO: 1308.

XX KW Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianemic; cytostatic; nootropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
 KW ds.

XX OS Homo sapiens.

XX PN WO200200928-A2.

XX PD 03-JAN-2002.

XX PF 02-JUL-2001; 2001WO-EP007537.

XX PR 30-JUN-2000; 2000DE-01032529.

XX PR 01-SEP-2000; 2000DE-01043826.

XX PA (EPIG-) EPIGENOMICS AG.

XX PI Olek A, Piepenbrock C, Berlin K;

XX DR WPI; 2002-130909/17.

XX PT Nucleic acid comprising fragment of chemically modified gene, useful for
 PT diagnosis and treatment of diseases associated with abnormal cytosine
 PT methylation.

XX PS Claim 1; SEQ ID NO 1308; 32pp + Sequence Listing; German.

XX CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention

XX SQ Sequence 19653 BP; 5918 A; 187 C; 4299 G; 9249 T; 0 U; 0 Other;

Query Match 84.8%; Score 17.8; DB 6; Length 19653;

Best Local Similarity 90.5%; Pred. No. 85;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTGTGTGGGAGTTATTGAGT 21

DB 17155 TTGTGTGGGAGTTATTGAGT 17175

RESULT 3

AAS73256

ID AAS73256 standard; cDNA; 730 BP.

XX AC AAS73256;

XX DT 13-FEB-2002 (first entry)

XX DE DNA encoding novel human diagnostic protein #9060.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US008631.

XX PR 31-MAR-2000; 2000US-00540217.

XX PR 23-AUG-2000; 2000US-00649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR P-FSDB; ABG09069.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.

XX PS Claim 1; SEQ ID NO 9060; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
 CC coding sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 730 BP; 220 A; 79 C; 245 G; 186 T; 0 U; 0 Other;

Query Match 80.0%; Score 16.8; DB 5; Length 730;

Best Local Similarity 90.0%; Pred. No. 1.8e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTGTGTGGGAGTTATTGAG 20
 Db 167 TTGTGTGGGAGTTATTGAG 186

RESULT 4
 ID AAS79261/c
 ID AAS79261 standard; cDNA; 2711 BP.

XX AC AAS79261;
 XX DT 13-FEB-2002 (first entry)
 XX DE

DE DNA encoding novel human diagnostic protein #15065.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 OS Homo sapiens.

XX WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US008631.

XX PR 31-MAR-2000; 2000US-00540217.

XX PR 23-AUG-2000; 2000US-00649167.

XX FA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR P-PSDB; ASG15074.

XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.

XX PS Claim 1; SEQ ID NO 15065; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligonucleotides, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
 CC coding sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pat_sequences

XX SQ Sequence 2711 BP; 1024 A; 662 C; 442 G; 583 T; 0 U; 0 Other;

Query Match 80.0%; Score 16.8; DB 5; Length 2711;
 Best Local Similarity 90.0%; Pred. No. 2.1e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTGTGTGGGAGTTATTGAG 20
 ||||| ||||| ||||| |||||

Db 2451 TTGTGTGGGAGTTATTGAG 2432

RESULT 5

ID AAT43351

XX AC AAT43351 standard; cDNA; 2764 BP.

XX AC AAT43351;

XX DT 08-SEP-1997 (first entry)

XX DE G-protein coupled receptor coding sequence.

XX KW G-protein coupled receptor; human; HIBEF51; transmembrane domain; ulcer;
 KW hormone; viral receptor; growth factor; neuroreceptor; neurotransmitter;
 KW signal transduction; central nervous system; hypertension; osteoporosis;
 KW angina pectoris; myocardial infarction; asthma; allergy; eating disorder;
 KW psychosis; depression; migraine; vomiting; stroke; cancer; hypotension;
 KW benign hypertrophy; Parkinson's disease; acute heart failure; therapy;
 KW urinary retention; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers
 XX CDS 832..1881
 XX FT /*tag= a
 XX FT /transl_except= seq: 1600..1602, aa: Thr

XX WO9639441-A1.

XX PD 12-DEC-1996.

XX PF 06-JUN-1995; 95WO-US007225.

XX PR 06-JUN-1995; 95WO-US007225.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Li Y, Adams MD;

XX DR WPI; 1997-043076/04.

XX DR P-PSDB; AAW08101.

XX PT Human G-protein coupled receptor, HIBEF51 - used to identify
 PT (ant)agonists, used in the treatment of asthma, angina pectoris,
 PT psychotic and neurological disorders, and eating disorders etc.

XX PS Claim 2; Page 46-48; 69pp; English.

XX CC This sequence represents the coding sequence for the human G-protein
 CC coupled receptor HIBEF51. The encoded protein is a 7-transmembrane domain
 CC receptor. G-protein coupled receptors include a wide range of
 CC biologically active receptors, such as hormone, viral, growth factor and
 CC neuroreceptors. Most of the receptors have conserved cysteine residues in
 CC the first two extracellular loops, which form disulphide bonds thought to
 CC stabilise the functional protein structure. Phosphorylation or lipidation
 CC of these residues can influence the signal transduction of some G-protein
 CC coupled receptors. G-protein coupled receptors are found at numerous
 CC sites within a mammalian host, and some are critical neurotransmitters in
 CC the central nervous system. Compounds that activate or inhibit the
 CC encoded receptor may be used for the treatment of patients which need to
 CC activate or inhibit a G-protein coupled receptor. Mutations in this
 CC sequence or the encoded protein may be identified by sequence analysis.
 CC Antagonists of the G-protein coupled receptor may be used for the
 CC treatment of hypertension, angina pectoris, myocardial infarction,
 CC ulcers, asthma, allergies, psychoses, cancer and benign hypertrophy. Agonists of
 CC stroke, eating disorders, cancer and benign hypertrophy. Agonists of
 CC encoded protein may be used in the treatment of Parkinson's disease,
 CC acute heart failure, hypotension, urinary retention and osteoporosis

XX SQ Sequence 2764 BP; 715 A; 587 C; 581 G; 881 T; 0 U; 0 Other;

Query Match 80.0%; Score 16.8; DB 2; Length 2764;

KW hypertension; hyperlipemia; heart disease; cancer; restenosis.
 XX Homo sapiens.
 OS
 PN WO2003048360-A1.
 XX
 XX 12-JUN-2003.
 PD
 XX
 PF 06-DEC-2002; 2002WO-JP012835.
 XX
 XX 07-DEC-2001; 2001JP-00375080.
 PR
 XX (NEWI-) NEW IND RES ORG.
 PA
 XX
 XX Yonezawa K, Hara K, Yoshino K, Tokunaga C;
 PI
 XX WPI; 2003-493616/46.
 DR
 XX Raptor protein which binds to a mammalian target of rapamycin, useful for
 PT the treatment, prevention and diagnosis of diabetes, hypertension and
 PT cancer and treatment of restenosis.
 PT
 XX
 XX Claim 7; SEQ ID NO 36; 272pp; Japanese.
 PS
 XX The invention relates to a novel human or mouse raptor protein or a
 CC sequence derived from these by addition, deletion and/or substitution of
 CC one or more amino acid residues and binds to a mammalian rapamycin target
 CC protein (mTOR) or to the mTOR signaling motif (TOS motif). mTOR is a
 CC kinase which regulates the phosphorylation of p70S6 kinase and eIF-4BP,
 CC which control cellular functions in response to signaling by receptors
 CC for insulin and amino acids respectively. The raptor proteins are useful
 CC for the treatment, prevention and diagnosis of diseases associated with
 CC the signaling pathways regulated by mTOR, such as diabetes, hypertension,
 CC hyperlipemia, heart disease, cancer and restenosis. The DNA encoding the
 CC raptor proteins is useful for the detection of pathological single
 CC nucleotide polymorphisms (SNP) in raptor. This sequence corresponds to a
 CC genomic fragment of the human mTOR sequence.
 XX
 SQ Sequence 50000 BP; 11873 A; 10980 C; 12056 G; 15090 T; 0 U; 1 Other;
 Query Match 80.0%; Score 16.8; DB 9; Length 50000;
 Best Local Similarity 90.0%; Pred. No. 2.9e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 OY 1 TTGTGTGGGAGTTATTCAG 20
 DB 8255 TTGTGTGGGAGTTACTGAG 8274
 RESULT 9
 ABX85261
 ID ABX85261 standard; DNA; 75899 BP.
 XX
 AC ABX85261;
 XX
 DT 13-AUG-2002 (first entry)
 DE Human genomic DNA for protein phosphatase 1B, PTP1B.
 XX
 XX Antisense; protein phosphatase 1B; PTP1B; ds; gene; human;
 KW type 2 diabetes; obesity; ovarian cancer; chronic myeloid leukaemia;
 KW hyperproliferative disease; antidiabetic; anorectic; cytostatic;
 KW blood glucose; gene therapy.
 XX
 OS Homo sapiens.
 XX
 XX US2002055479-A1.
 PN
 XX
 PD 09-MAY-2002.
 XX
 PF 14-MAY-2001; 2001US-00854883.
 XX
 XX 18-JAN-2000; 2000US-00487368.
 PR

PR 31-JUL-2000; 2000US-00629644.
 XX
 XX (COWS/) COWSERT L M.
 PA (WYAT/) WYATT J.
 PA (FREI/) FREIER S M.
 PA (MONI/) MONIA B P.
 PA (BUTL/) BUTLER M M.
 PA (MCKA/) MCKAY R.
 XX
 XX Cowsert LM, Wyatt J, Freier SM, Monia BP, Butler MM, McKay R;
 PI WPI; 2002-462914/49.
 DR
 XX
 XX Compound for inhibiting the expression of protein phosphatase 1B (PTP1B)
 PT and for treating diabetes, cancer, or obesity, comprises an antisense
 PT oligonucleotide targeted to nucleic acid encoding PTP1B.
 PT
 XX
 XX Example 22; Page 75-108; 133pp; English.
 PS
 XX The invention relates to a compound of 8-50 nucleobases in length
 CC targeted to a nucleic acid encoding protein phosphatase 1B (PTP1B), where
 CC the compound specifically hybridises with and inhibits the expression of
 CC PTP1B (e.g. an antisense oligonucleotide). Also included are (1) a
 CC compound of 8-50 nucleobases in length which specifically hybridises with
 CC an 8 nucleobase portion of an active site on a nucleic acid encoding
 CC PTP1B; (2) inhibiting the expression of PTP1B in cells or tissues
 CC comprising contacting the cells or tissues with the compound; treating an
 CC animal having or suspected of having a disease or condition associated
 CC with PTP1B comprising administering the compound; (4) decreasing blood
 CC sugar levels in an animal comprising administering the compound; (5)
 CC preventing or delaying the onset of a disease or condition associated
 CC with PTP1B in an animal comprising administering the compound; and (6)
 CC preventing or delaying the onset of an increase in blood glucose levels
 CC in an animal comprising administering the compound. The compound is used
 CC to inhibit the expression of PTP1B in cells or tissues, to treat or
 CC prevent or delay the onset of a disease or condition associated with
 CC PTP1B, such as type 2 diabetes, obesity, cancer (especially ovarian
 CC cancer, chronic myeloid leukaemia and hyperproliferative diseases in an
 CC animal having or suspected of having the disease or condition, and for
 CC decreasing blood sugar levels or preventing or delaying the onset of an
 CC increase in blood glucose levels in an animal. The compound is also used
 CC in diagnostics, therapeutics, prophylaxis, and in research reagents and
 CC kits. The present sequence is human PTP1B genomic DNA
 XX
 SQ Sequence 75899 BP; 19300 A; 16420 C; 16761 G; 23418 T; 0 U; 0 Other;
 Query Match 80.0%; Score 16.8; DB 6; Length 75899;
 Best Local Similarity 90.0%; Pred. No. 3.1e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 OY 1 TTGTGTGGGAGTTATTCAG 20
 DB 59270 TTGTGTGGGAGTTATTCAG 59289
 RESULT 10
 ADE56603
 ID ADE56603 standard; DNA; 1932 BP.
 XX
 AC ADE56603;
 XX
 XX 29-JAN-2004 (first entry)
 DT
 DE Human gene X56351, SEQ ID NO 2457.
 XX
 XX Human; ds; gene; pain; neuronal tissue; gene therapy;
 KW spinal segmental nerve injury; chronic constriction injury; CCI;
 KW spared nerve injury; SNI; Chung.
 XX
 OS Homo sapiens.
 XX
 PN WO2003016475-A2.
 PR

PD 27-FEB-2003.
 XX
 PF 14-AUG-2002; 2002WO-US025765.
 XX
 PR 14-AUG-2001; 2001US-0312147P.
 XX
 PR 01-NOV-2001; 2001US-0346382P.
 PR 26-NOV-2001; 2001US-0333347P.
 XX
 XX (GEOH) GEN HOSPITAL CORP.
 PA (FARB) BAYER AG.
 PA
 PI Woolf C, D'urso D, Befort K, Costigan M;
 XX
 XX WPI; 2003-268312/26.
 DR GENBANK; X56351.
 PS
 XX New composition comprising two or more isolated polypeptides, useful for
 PT preparing a medicament for treating pain in an animal.
 XX
 XX Claim 1; Page; 1017pp; English.
 XX
 CC The invention discloses a composition comprising two or more isolated rat
 CC or human polynucleotides or a polynucleotide which represents a fragment,
 CC derivative or allelic variation of the nucleic acid sequence. Also
 CC claimed are a vector comprising the novel polynucleotide, a host cell
 CC comprising the vector, a method for identifying a nucleotide sequence
 CC which is differentially regulated in an animal subjected to pain and a
 CC kit to perform the method, an array, a method for identifying an agent
 CC that increases or decreases the expression of the polynucleotide sequence
 CC that is differentially expressed in neuronal tissue of a first animal
 CC subjected to pain, a method for identifying a compound which regulates
 CC the expression of a polynucleotide sequence which is differentially
 CC expressed in an animal subjected to pain, a method for identifying a
 CC compound that regulates the activity of one or more of the
 CC polynucleotides, a method for producing a pharmaceutical composition, a
 CC method for identifying a compound or small molecule that regulates the
 CC activity in an animal of one or more of the polypeptides given in the
 CC specification, a method for identifying a compound useful in treating
 CC pain and a pharmaceutical composition comprising the one or more
 CC polypeptides or their antibodies. The polynucleotide or the compound that
 CC modulates its activity is useful for preparing a medicament for treating
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
 CC therapy). The sequence presented is a human DNA (shown in Table 2 of the
 CC specification) which encodes one of the polypeptides of the invention
 CC which is differentially expressed during pain. Note: The sequence data
 CC for this patent did not form part of the printed specification, but was
 CC obtained in electronic form directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1932 BP; 500 A; 495 C; 477 G; 460 T; 0 U; 0 Other;
 Query Match 78.1%; Score 16.4; DB 9; Length 1932;
 Best Local Similarity 94.4%; Pred. No. 3.1e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 2 TGTGTGGGGAGTTATGGA 19
 DB 775 TGTGTGGGGAGTTATGGA 792
 RESULT 11
 ADE56607
 ID ADE56607 standard; DNA; 1932 BP.
 XX
 AC ADE56607;
 XX
 DT 29-JAN-2004 (first entry).
 XX
 DE Human gene X56351, SEQ ID NO 2461.
 DE
 XX Human; ds; gene; pain; neuronal tissue; gene therapy;
 KW spinal segmental nerve injury; chronic constriction injury; CCI;
 KW

KW spared nerve injury; SNI; Chung.
 XX Homo sapiens.
 XX WO2003016475-A2.
 XX 27-FEB-2003.
 PD
 XX 14-AUG-2002; 2002WO-US025765.
 PF
 XX 14-AUG-2001; 2001US-0312147P.
 PR
 PR 01-NOV-2001; 2001US-0346382P.
 PR 26-NOV-2001; 2001US-0333347P.
 XX
 XX (GEOH) GEN HOSPITAL CORP.
 PA (FARB) BAYER AG.
 PA
 PI Woolf C, D'urso D, Befort K, Costigan M;
 XX
 XX WPI; 2003-268312/26.
 DR GENBANK; X56351.
 PS
 XX New composition comprising two or more isolated polypeptides, useful for
 PT preparing a medicament for treating pain in an animal.
 XX
 XX Claim 1; Page; 1017pp; English.
 XX
 CC The invention discloses a composition comprising two or more isolated rat
 CC or human polynucleotides or a polynucleotide which represents a fragment,
 CC derivative or allelic variation of the nucleic acid sequence. Also
 CC claimed are a vector comprising the novel polynucleotide, a host cell
 CC comprising the vector, a method for identifying a nucleotide sequence
 CC which is differentially regulated in an animal subjected to pain and a
 CC kit to perform the method, an array, a method for identifying an agent
 CC that increases or decreases the expression of the polynucleotide sequence
 CC that is differentially expressed in neuronal tissue of a first animal
 CC subjected to pain, a method for identifying a compound which regulates
 CC the expression of a polynucleotide sequence which is differentially
 CC expressed in an animal subjected to pain, a method for identifying a
 CC compound that regulates the activity of one or more of the
 CC polynucleotides, a method for producing a pharmaceutical composition, a
 CC method for identifying a compound or small molecule that regulates the
 CC activity in an animal of one or more of the polypeptides given in the
 CC specification, a method for identifying a compound useful in treating
 CC pain and a pharmaceutical composition comprising the one or more
 CC polypeptides or their antibodies. The polynucleotide or the compound that
 CC modulates its activity is useful for preparing a medicament for treating
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
 CC therapy). The sequence presented is a human DNA (shown in Table 2 of the
 CC specification) which encodes one of the polypeptides of the invention
 CC which is differentially expressed during pain. Note: The sequence data
 CC for this patent did not form part of the printed specification, but was
 CC obtained in electronic form directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1932 BP; 500 A; 495 C; 477 G; 460 T; 0 U; 0 Other;
 Query Match 78.1%; Score 16.4; DB 9; Length 1932;
 Best Local Similarity 94.4%; Pred. No. 3.1e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 2 TGTGTGGGGAGTTATGGA 19
 DB 775 TGTGTGGGGAGTTATGGA 792
 RESULT 12
 ADE56611
 ID ADE56611 standard; DNA; 1932 BP.
 XX
 AC ADE56611;
 XX

DT 29-JAN-2004 (first entry)
XX Human gene X56351, SEQ ID NO 2465.
XX
XX
XX Human; ds; gene; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; chronic constriction injury; CCI;
KW spared nerve injury; SNI; Chung.
XX
XX Homo sapiens.
XX
XX WO2003016475-A2.
XX
XX 27-FEB-2003.
XX
XX 14-AUG-2002; 2002WO-US025765.
XX
XX 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX
XX (GEHO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.
XX
XX Woolf C, D'urso D, Befort K, Costigan M;
XX
XX WPI; 2003-268312/26.
XX GENBANK; X56351.
XX
XX New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX
XX Claim 1; Page; 1017pp; English.
XX
XX The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a human DNA (shown in Table 2 of the
CC specification) which encodes one of the polypeptides of the invention
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 1932 BP; 500 A; 495 C; 477 G; 460 T; 0 U; 0 Other;
Query Match 78.1%; Score 16.4; DB 9; Length 1932;
Best Local Similarity 94.4%; Pred. No. 3.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT 13
ADE56599
ID ADE56599 standard; DNA; 1932 BP.
XX
XX ADE56599;
AC
XX 29-JAN-2004 (first entry)
DT
XX Human gene X56351, SEQ ID NO 2453.
DE
XX Human; ds; gene; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; chronic constriction injury; CCI;
KW spared nerve injury; SNI; Chung.
XX
XX Homo sapiens.
OS
XX WO2003016475-A2.
PN
XX 27-FEB-2003.
PD
XX
XX 14-AUG-2002; 2002WO-US025765.
PF
XX 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX
XX (GEHO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.
XX
XX Woolf C, D'urso D, Befort K, Costigan M;
XX
XX WPI; 2003-268312/26.
XX GENBANK; X56351.
XX
XX New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX
XX Claim 1; Page; 1017pp; English.
XX
XX The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a human DNA (shown in Table 2 of the
CC specification) which encodes one of the polypeptides of the invention
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 1932 BP; 500 A; 495 C; 477 G; 460 T; 0 U; 0 Other;
Query Match 78.1%; Score 16.4; DB 9; Length 1932;
Best Local Similarity 94.4%; Pred. No. 3.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TGTGGGGAGTTATTGA 19
 DB 775 TGTGGGGAGTTATTGA 792

RESULT 14
 ABQ66985
 ID ABQ66985 standard; DNA; 5259 BP.
 XX
 AC ABQ66985;
 XX
 DT 28-AUG-2002 (first entry)
 XX
 DE Human angiogenesis associated polynucleotide SEQ ID NO 15.
 XX
 KW Human; angiogenesis; methylation; eye disease; glaucoma; tumour;
 KW inflammation; rheumatoid arthritis; diabetic retinopathy; antitubercu;
 KW macular degeneration; inflammatory bowel disease; Crohn's disease;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antiarteriosclerotic; ds.
 XX
 OS Homo sapiens.
 XX
 FN WO200246454-A2.
 XX
 PD 13-JUN-2002.
 XX
 PF 06-DEC-2001; 2001WO-EP014320.
 XX
 PR 06-DEC-2000; 2000DE-01061338.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Schacht O;
 XX
 DR WPI; 2002-500450/53.
 XX
 PT New nucleic acid fragments from chemically treated angiogenesis-
 PT associated genes, useful for determining methylation status, e.g. in
 PT diagnosis or treatment of cancer.
 XX
 PS Claim 1; SEQ ID NO 15; 41pp + Sequence Listing; German.
 XX
 CC The invention relates to a nucleic acid (I) comprising a segment of 18
 CC bases of chemically pretreated DNA of angiogenesis-associated genes (II)
 CC having sequences (ABQ66971-ABQ67178) or their complements. (I), also
 CC related oligomers, are used to evaluate the methylation status and/or
 CC single-nucleotide polymorphisms, in angiogenesis-related genes, for
 CC diagnosis and treatment of eye diseases, proliferative retinopathy,
 CC neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,
 CC diabetic retinopathy, macular degeneration caused by neovascularisation,
 CC psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and
 CC Crohn's disease. Note: The sequence data for this patent did not form
 CC part of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 5259 BP; 1146 A; 128 C; 1518 G; 2467 T; 0 U; 0 Other;

Query Match 78.1%; Score 16.4; DB 6; Length 5259;
 Best Local Similarity 94.4%; Fred. No. 3.5e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GTGTGGGGAGTTATTGAG 20
 DB 1596 GTGTGGGGAGTTATTGGG 1613

RESULT 15
 ABL22880/c
 ID ABL22880 standard; DNA; 8410 BP.
 XX
 AC ABL22880;

XX 26-MAR-2002 (first entry)
 DT
 XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 20113.
 DE
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.
 KW
 XX Drosophila melanogaster.
 OS
 XX WO200171042-A2.
 PN
 XX 27-SEP-2001.
 PD
 XX 23-MAR-2001; 2001WO-US009231.
 PF
 XX 23-MAR-2000; 2000US-0191637P.
 PR
 PR 11-JUL-2000; 2000US-00614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI; 2001-656860/75.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 PT
 XX Claim 1; SEQ ID NO 20113; 21pp + Sequence Listing; English.
 PS
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 8410 BP; 2048 A; 1870 C; 2088 G; 2404 T; 0 U; 0 Other;

Query Match 78.1%; Score 16.4; DB 4; Length 8410;
 Best Local Similarity 94.4%; Fred. No. 3.7e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 TGTGGGGAGTTATTGAGT 21
 DB 230 TGTGGGGAGTTATTGAGT 213

Search completed: April 24, 2004, 18:44:59
 Job time : 180.593 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2004, 18:26:15 ; Search time 38.889 Seconds
(without alignments)
299.673 Million cell updates/sec

Title: US-10-084-555a-115
Perfect score: 21
Sequence: 1 ttgtgtgggagttattgag 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16.8	80.0	2764	2	US-08-465-971B-1
2	16.2	77.1	4937	1	US-08-038-682-3
3	16.2	77.1	4937	1	US-08-302-832-3
4	16.2	77.1	4937	2	US-08-530-198-3
5	16.2	77.1	4937	2	US-08-469-880-3
6	16.2	77.1	4937	2	US-08-728-470-3
7	16.2	77.1	4937	2	US-08-617-697-3
8	16.2	77.1	4937	3	US-08-719-641-3
9	16.2	77.1	4937	4	US-09-206-942-70
10	16.2	77.1	5116	1	US-08-038-682-1
11	16.2	77.1	5116	1	US-08-302-832-1
12	16.2	77.1	5116	2	US-08-530-198-1
13	16.2	77.1	5116	2	US-08-469-880-1
14	16.2	77.1	5116	2	US-08-728-470-1
15	16.2	77.1	5116	2	US-08-617-697-1
16	16.2	77.1	5116	3	US-08-719-641-1
17	16.2	77.1	5116	4	US-09-206-942-66
18	16.2	77.1	9171	1	US-08-038-682-5
19	16.2	77.1	9171	1	US-08-302-832-5
20	16.2	77.1	9171	2	US-08-530-198-5
21	16.2	77.1	9171	2	US-08-469-880-5
22	16.2	77.1	9171	2	US-08-728-470-5
23	16.2	77.1	9171	2	US-08-617-697-5
24	16.2	77.1	9171	3	US-08-719-641-5
25	16.2	77.1	9323	1	US-08-038-682-6
26	16.2	77.1	9323	1	US-08-302-832-6
27	16.2	77.1	9323	2	US-08-530-198-6

28	16.2	77.1	9323	2	US-08-469-880-6	Sequence 6, Appli
29	16.2	77.1	9323	2	US-08-728-470-6	Sequence 6, Appli
30	16.2	77.1	9323	2	US-08-617-697-6	Sequence 6, Appli
31	16.2	77.1	9323	3	US-08-719-641-6	Sequence 6, Appli
C 32	16	76.2	681	4	US-09-149-476-189	Sequence 189, Appl
33	15.8	75.2	276	4	US-09-313-294A-2823	Sequence 2823, Ap
34	15.8	75.2	193303	4	US-09-497-855A-37	Sequence 37, Appl
35	15.8	75.2	193303	4	US-09-497-855A-44	Sequence 44, Appl
C 36	15.2	72.4	1956	4	US-09-016-434-136	Sequence 136, Appl
C 37	15.2	72.4	2934	3	US-09-149-934-2	Sequence 2, Appli
C 38	15.2	72.4	3032	4	US-09-833-381-1482	Sequence 1482, Ap
C 39	15.2	72.4	3032	4	US-09-833-381-1483	Sequence 1483, Ap
40	15.2	72.4	5844	4	US-10-204-708-90	Sequence 90, Appli
41	15.2	72.4	90050	3	US-09-245-041-5	Sequence 5, Appli
C 42	15.2	72.4	392000	4	US-10-027-983-11	Sequence 11, Appl
C 43	15	71.4	513	4	US-09-115-407-37	Sequence 37, Appl
C 44	14.8	70.5	72	1	US-08-303-275-89	Sequence 89, Appl
C 45	14.8	70.5	72	2	US-08-658-665-79	Sequence 79, Appl

ALIGNMENTS

RESULT 1
US-08-465-971B-1
; Sequence 1, Application US/08465971B
; Patent No. 5942414
; GENERAL INFORMATION:
; APPLICANT: Yi Li and Mark D. Adams
; TITLE OF INVENTION: Human G-Protein Receptor HIBEP51
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan,
; ADDRESSEE: Cecchi, Stewart & Olstein
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: NJ
; COUNTRY: US
; ZIP: 07068-1739
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,971B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: MULLINS, J.G.
; REGISTRATION NUMBER: 33073
; REFERENCE/DOCKET NUMBER: 325800-453 (PF187)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2764 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-465-971B-1

Query Match 80.0%; Score 16.8; DB 2; Length 2764;
Best Local Similarity 90.0%; Pred. No. 17;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTGTGTGGGAGTTATTGAG 20
|||||
Db 289 TTGTGTGGGAGTTATTGAG 308

```
RESULT 2
US-08-038-682-3
; Sequence 3, Application US/08038682
; Patent No. 5549897
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GEME III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/038,682
; FILING DATE: 16-MAR-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-293
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4937 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-038-682-3
Query Match 77.1%; Score 16.2; DB 1; Length 4937;
Best Local Similarity 85.7%; Pred. No. 39;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TTGTGTGGGAGTTATTGAGT 21
|||
Db 1523 TTGTGTGGGCGATATTGCGT 1543

RESULT 3
US-08-038-682-3
; Sequence 3, Application US/08038682
; Patent No. 5549897
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: Of No. 5603938-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/302,832
; FILING DATE: 16-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; APPLICATION NUMBER: US pct/us93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-404
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4937 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-302-832-3
Query Match 77.1%; Score 16.2; DB 1; Length 4937;
Best Local Similarity 85.7%; Pred. No. 39;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TTGTGTGGGAGTTATTGAGT 21
|||
Db 1523 TTGTGTGGGCGATATTGCGT 1543

RESULT 4
US-08-530-198-3
; Sequence 3, Application US/08530198
; Patent No. 5865065
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GEME III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/530,198
; FILING DATE: 13-DEC-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: JWB-1186
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4937 base pairs
```

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-530-198-3

Query Match 77.1%; Score 16.2; DB 2; Length 4937;
Best Local Similarity 85.7%; Pred. No. 39;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TTGTGTGGGGAGTTATTGAGT 21
|||||
DB 1523 TTGTGTGGGGCGATATTGCGT 1543

RESULT 5
US-08-469-880-3
Sequence 3, Application US/08469880
Patent No. 5876733
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J.
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5876733-Typeable Haemophilus
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469.880
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-516 MIS:vg
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 4937 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-469-880-3

Query Match 77.1%; Score 16.2; DB 2; Length 4937;
Best Local Similarity 85.7%; Pred. No. 39;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TTGTGTGGGGAGTTATTGAGT 21
|||||
DB 1523 TTGTGTGGGGCGATATTGCGT 1543

RESULT 6
US-08-728-470-3
Sequence 3, Application US/08728470
Patent No. 5928651
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J.
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728.470
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-633
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 4937 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-728-470-3

Query Match 77.1%; Score 16.2; DB 2; Length 4937;
Best Local Similarity 85.7%; Pred. No. 39;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TTGTGTGGGGAGTTATTGAGT 21
|||||
DB 1523 TTGTGTGGGGCGATATTGCGT 1543

RESULT 7
US-08-617-697-3
Sequence 3, Application US/08617697
Patent No. 5977336
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J.
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.

;; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
;; NUMBER OF SEQUENCES: 8
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Shoemaker and Mattare, Ltd
;; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;; Bldg. 1
;; City: Arlington
;; STATE: Virginia
;; COUNTRY: U.S.A.
;; ZIP: 22202-0286
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/038,682
;; FILING DATE: 16-MAR-1993
;; CLASSIFICATION: 424
;; ATTORNEY/AGENT INFORMATION:
;; NAME: BERKSTRESSER, JERRY W
;; REGISTRATION NUMBER: 22,651
;; REFERENCE/DOCKET NUMBER: 1038-293
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (703) 415-0810
;; TELEFAX: (703) 415-0813
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 5116 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; US-08-038-682-1
;;
Query Match 77.1%; Score 16.2; DB 1; Length 5116;
Best Local Similarity 85.7%; Pred. No. 39;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 TTGTGTGGGGAGTTATTGAGT 21
Db 1522 TTGTGTGGGGAGTTATTGCGT 1542
RESULT 11
US-08-832-832-1
;; Sequence 1, Application US/08302832
;; Patent No. 5603938
;; GENERAL INFORMATION:
;; APPLICANT: Barenkamp, Stephen J
;; TITLE OF INVENTION: High Molecular Weight Surface Proteins
;; TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
;; NUMBER OF SEQUENCES: 8
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Shoemaker and Mattare, Ltd.
;; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;; Bldg. 1
;; City: Arlington
;; STATE: Virginia
;; COUNTRY: U.S.A.
;; ZIP: 22202-0286
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/302,832
;; FILING DATE: 16-SEP-1994
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: GB 9205704.1
;; FILING DATE: 16-MAR-1992

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US pct/us93/02166
;; FILING DATE: 16-MAR-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Berkstresser, Jerry W
;; REGISTRATION NUMBER: 22,651
;; REFERENCE/DOCKET NUMBER: 1038-404
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (703) 415-0810
;; TELEFAX: (703) 415-0813
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 5116 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; US-08-302-832-1
;;
Query Match 77.1%; Score 16.2; DB 1; Length 5116;
Best Local Similarity 85.7%; Pred. No. 39;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 TTGTGTGGGGAGTTATTGAGT 21
Db 1522 TTGTGTGGGGAGTTATTGCGT 1542
RESULT 12
US-08-530-198-1
;; Sequence 1, Application US/08530198
;; Patent No. 5869085
;; GENERAL INFORMATION:
;; APPLICANT: BARENKAMP, STEPHEN J
;; APPLICANT: ST. GEME III, JOSEPH W
;; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
;; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
;; NUMBER OF SEQUENCES: 8
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Shoemaker and Mattare, Ltd
;; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;; Bldg. 1
;; City: Arlington
;; STATE: Virginia
;; COUNTRY: U.S.A.
;; ZIP: 22202-0286
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/530,198
;; FILING DATE: 13-DEC-1995
;; CLASSIFICATION: 424
;; ATTORNEY/AGENT INFORMATION:
;; NAME: BERKSTRESSER, JERRY W
;; REGISTRATION NUMBER: 22,651
;; REFERENCE/DOCKET NUMBER: JWB-1186
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (703) 415-0810
;; TELEFAX: (703) 415-0813
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 5116 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; US-08-530-198-1
;;
Query Match 77.1%; Score 16.2; DB 2; Length 5116;
Best Local Similarity 85.7%; Pred. No. 39;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TTGTGTGGGGAGTTATTGAGT 21
|||||
Db 1522 TTGTGTGGGGCGATATTCGT 1542

RESULT 13

US-08-469-880-1
; Sequence 1, Application US/08469880
; Patent No. 5876733
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5876733-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,880
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstreser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-516 MIS.vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5116 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-469-880-1

Query Match 77.1%; Score 16.2; DB 2; Length 5116;
Best Local Similarity 85.7%; Pred. No. 39;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TTGTGTGGGGAGTTATTGAGT 21
|||||
Db 1522 TTGTGTGGGGCGATATTCGT 1542

RESULT 14

US-08-728-470-1
; Sequence 1, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J

; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,470
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstreser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-633
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5116 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-728-470-1

Query Match 77.1%; Score 16.2; DB 2; Length 5116;
Best Local Similarity 85.7%; Pred. No. 39;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TTGTGTGGGGAGTTATTGAGT 21
|||||
Db 1522 TTGTGTGGGGCGATATTCGT 1542

RESULT 15

US-08-617-697-1
; Sequence 1, Application US/08617697
; Patent No. 5977336
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/617,697
FILING DATE: 01-APR-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 05-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-557
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5116 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-617-697-1

Query Match 77.1%; Score 16.2; DB 2; Length 5116;
Best Local Similarity 85.7%; Pred.No.39;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TTGTGGGGAGTTATTGAGT 21
|||||
DB 1522 TTGTGGGGGATATTGCGT 1542

Search completed: April 24, 2004, 21:13:33
Job time : 40.889 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2004, 19:25:41 ; Search time 162.296 Seconds
(without alignments)
583.385 Million cell updates/sec

Title: US-10-084-555A-115

Perfect score: 21

Sequence: 1 ttgtgtggggaggtattgagt 21

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2907579 seqs, 2254313464 residues

Total number of hits satisfying chosen parameters: 5815158

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

1:	/cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:
2:	/cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:
3:	/cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:
4:	/cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:
5:	/cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:
6:	/cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:
7:	/cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:
8:	/cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:
9:	/cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:
10:	/cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:
11:	/cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:
12:	/cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
13:	/cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:
14:	/cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:
15:	/cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:
16:	/cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:
17:	/cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:
18:	/cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:
19:	/cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17.8	84.8	13653	US-10-311-455-1308	Sequence 1308, Ap
2	16.8	80.0	607	US-10-027-632-115705	Sequence 115705,
3	16.8	80.0	607	US-10-027-632-115706	Sequence 115706,
4	16.8	80.0	607	US-10-027-632-115705	Sequence 115705,
5	16.8	80.0	607	US-10-027-632-115706	Sequence 115706,
6	16.8	80.0	2764	US-10-006-394-1	Sequence 1, Appli
7	16.8	80.0	5557	US-10-311-455-1520	Sequence 1520, Ap
8	16.8	80.0	5756	US-10-311-455-559	Sequence 559, App
9	16.8	80.0	75899	US-09-854-883-243	Sequence 243, App
10	16.8	80.0	75899	US-10-360-510-243	Sequence 243, App
11	16.8	80.0	43982	US-10-087-192-454	Sequence 454, App
12	16.2	77.1	388	US-10-424-599-85816	Sequence 85816, A
13	16.2	77.1	672	US-10-027-632-185265	Sequence 185265,
14	16.2	77.1	672	US-10-027-632-185265	Sequence 185265,

c 15	16.2	77.1	703	13	US-10-027-632-239897	Sequence 239897,
c 16	16.2	77.1	703	16	US-10-027-632-239897	Sequence 239897,
c 17	16.2	77.1	1339	13	US-10-027-632-263908	Sequence 263908,
18	16.2	77.1	1339	16	US-10-027-632-263908	Sequence 263908,
c 19	16.2	77.1	1369	13	US-10-027-632-123600	Sequence 123600,
c 20	16.2	77.1	1369	16	US-10-027-632-123600	Sequence 123600,
21	16.2	77.1	2400	13	US-10-027-632-264395	Sequence 264395,
22	16.2	77.1	2400	16	US-10-027-632-264395	Sequence 264395,
23	16.2	77.1	4937	14	US-10-092-880-3	Sequence 3, Appli
24	16.2	77.1	4937	15	US-10-193-764-66	Sequence 66, Appli
25	16.2	77.1	5070	15	US-10-311-455-774	Sequence 774, App
26	16.2	77.1	5116	14	US-10-092-880-1	Sequence 1, Appli
27	16.2	77.1	5116	15	US-10-193-764-62	Sequence 62, Appli
28	16.2	77.1	5393	13	US-10-058-270A-71	Sequence 71, Appli
29	16.2	77.1	5488	15	US-10-311-455-1429	Sequence 1429, Ap
30	16.2	77.1	5814	15	US-10-311-455-1534	Sequence 1534, Ap
31	16.2	77.1	6063	15	US-10-240-453-268	Sequence 268, App
32	16.2	77.1	6503	15	US-10-311-455-693	Sequence 693, App
33	16.2	77.1	9007	13	US-10-221-714A-337	Sequence 337, App
34	16.2	77.1	9171	14	US-10-092-880-5	Sequence 5, Appli
35	16.2	77.1	9323	14	US-10-092-880-6	Sequence 6, Appli
36	16.2	77.1	9717	9	US-09-764-847-1581	Sequence 1581, Ap
c 37	16.2	77.1	9717	10	US-09-764-847-1581	Sequence 1581, Ap
38	16.2	77.1	9717	15	US-10-092-154-1581	Sequence 336, App
39	16.2	77.1	15674	15	US-10-311-455-336	Sequence 30, Appli
40	16.2	77.1	15674	15	US-10-240-485-30	Sequence 995, App
41	16.2	77.1	16236	15	US-10-311-455-995	Sequence 54, Appli
42	16.2	77.1	17421	15	US-10-239-676-54	Sequence 54, Appli
43	16.2	77.1	17421	15	US-10-240-453-56	Sequence 46, App
44	16.2	77.1	24259	13	US-10-221-714A-416	Sequence 416, App
45	16.2	77.1	32607	13	US-10-087-192-106	Sequence 106, App

ALIGNMENTS

RESULT 1
US-10-311-455-1308
; Sequence 1308, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1308
; LENGTH: 19653
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1308

Query Match 84.8%; Score 17.8; DB 15; Length 19653;
Best Local Similarity 90.5%; Pred. No. 95;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 TTGTGTGGGGAGGTATTGAGT 21
DB 17155 TGGTGTGGGGAGGTATTAGAGT 17175

```
RESULT 2
US-10-027-632-115705
; Sequence 115705, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 115705
; LENGTH: 607
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-115705

Query Match      80.0%; Score 16.8; DB 13; Length 607;
Best Local Similarity 90.0%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTGTGTGGGGAGTTATTGAG 20
Db 50 TTGTGTGGAGGAGTAATTGAG 69

RESULT 3
US-10-027-632-115706
; Sequence 115706, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 115706
; LENGTH: 607
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-115706

Query Match      80.0%; Score 16.8; DB 13; Length 607;
Best Local Similarity 90.0%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTGTGTGGGGAGTTATTGAG 20
Db 50 TTGTGTGGAGGAGTAATTGAG 69

RESULT 5
US-10-027-632-115706
; Sequence 115706, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
```

```
US-10-027-632-115706
Query Match      80.0%; Score 16.8; DB 13; Length 607;
Best Local Similarity 90.0%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTGTGTGGGGAGTTATTGAG 20
Db 50 TTGTGTGGAGGAGTAATTGAG 69

RESULT 4
US-10-027-632-115705
; Sequence 115705, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 115705
; LENGTH: 607
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-115705

Query Match      80.0%; Score 16.8; DB 16; Length 607;
Best Local Similarity 90.0%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTGTGTGGGGAGTTATTGAG 20
Db 50 TTGTGTGGAGGAGTAATTGAG 69

RESULT 5
US-10-027-632-115706
; Sequence 115706, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
```

```
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 115706
; LENGTH: 607
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-115706

Query Match      80.0%; Score 16.8; DB 16; Length 607;
Best Local Similarity 90.0%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TTGTGTGGGGAGTTATTGAG 20
      ||||||| ||||| ||||| ||||| |||||
DB      50 TTGTGTGGAGGAGTAATTGAG 69

RESULT 6
US-10-006-394-1
; Sequence 1, Application US/10006394
; Publication No. US20020086365A1
; GENERAL INFORMATION:
; APPLICANT: Yi Li and Mark D. Adams
; TITLE OF INVENTION: Human G-Protein Receptor HIBF51
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan,
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: NJ
; COUNTRY: US
; ZIP: 07068-1739
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/006,394
; FILING DATE: 10-Dec-2001
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/228,420
; FILING DATE: <unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: MULLINS, J.G.
; REGISTRATION NUMBER: 33073
; REFERENCE/DOCKET NUMBER: 325800-453 (PF187)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2764 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-006-394-1

Query Match      80.0%; Score 16.8; DB 13; Length 2764;
Best Local Similarity 90.0%; Pred. No. 2.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TTGTGTGGGGAGTTATTGAG 20
      ||||||| ||||| ||||| ||||| |||||
DB      50 TTGTGTGGAGGAGTAATTGAG 69

; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 115706
; LENGTH: 607
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-115706

Query Match      80.0%; Score 16.8; DB 16; Length 607;
Best Local Similarity 90.0%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TTGTGTGGGGAGTTATTGAG 20
      ||||||| ||||| ||||| ||||| |||||
DB      50 TTGTGTGGAGGAGTAATTGAG 69

RESULT 7
US-10-311-455-1520
; Sequence 1520, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Deter
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1520
; LENGTH: 5557
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1520

Query Match      80.0%; Score 16.8; DB 15; Length 5557;
Best Local Similarity 90.0%; Pred. No. 2.6e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TTGTGTGGGGAGTTATTGAG 20
      ||||||| ||||| ||||| ||||| |||||
DB      4074 TAGTGTGGGAGTATTGGG 4093

RESULT 8
US-10-311-455-559
; Sequence 559, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Deter
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 559
; LENGTH: 5756
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-559

Query Match      80.0%; Score 16.8; DB 15; Length 5756;
Best Local Similarity 90.0%; Pred. No. 2.6e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TTGTGTGGGGAGTTATTGAG 20
      ||||||| ||||| ||||| ||||| |||||
```

Db 3316 TTGTGTAGGAGTTATTGAG 3335
||||| |||||||

RESULT 9
US-09-854-883-243
; Sequence 243, Application US/09854883
; Patent No. US20020055479A1
; GENERAL INFORMATION:
; APPLICANT: Lex M. Cowsett
; APPLICANT: Jacqueline Wyatt
; APPLICANT: Susan M. Freier
; APPLICANT: Brett P. Monia
; APPLICANT: Madeline M. Butler
; APPLICANT: Robert McKay
; TITLE OF INVENTION: ANTISENSE MODULATION OF PTP1B EXPRESSION
; FILE REFERENCE: ISPH-0576
; CURRENT APPLICATION NUMBER: US/09/854,883
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 09/629,644
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 09/487,368
; PRIOR FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 389
; SEQ ID NO 243
; LENGTH: 75899
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-854-883-243

Query Match 80.0%; Score 16.8; DB 9; Length 75899;
Best Local Similarity 90.0%; Pred. No. 3.3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTGTGTGGGAGTTATTGAG 20
||||| |||||||

Db 59270 TTGTGTGGGAGTTATTGAG 59289

RESULT 10
US-10-360-510-243
; Sequence 243, Application US/10360510
; Publication No. US20030220282A1
; GENERAL INFORMATION:
; APPLICANT: Lex M. Cowsett
; APPLICANT: Jacqueline Wyatt
; APPLICANT: Susan M. Freier
; APPLICANT: Brett P. Monia
; APPLICANT: Madeline M. Butler
; APPLICANT: Robert McKay
; TITLE OF INVENTION: ANTISENSE MODULATION OF PTP1B EXPRESSION
; FILE REFERENCE: ISPH-0576
; CURRENT APPLICATION NUMBER: US/10/360,510
; CURRENT FILING DATE: 2003-02-07
; PRIOR APPLICATION NUMBER: US/09/854,883
; PRIOR FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 09/629,644
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 09/487,368
; PRIOR FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 389
; SEQ ID NO 243
; LENGTH: 75899
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-10-360-510-243

Query Match 80.0%; Score 16.8; DB 16; Length 75899;
Best Local Similarity 90.0%; Pred. No. 3.3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTGTGTGGGAGTTATTGAG 20
||||| |||||||

Db 59270 TTGTGTGGGAGTTATTGAG 59289

RESULT 11
US-10-087-192-454
; Sequence 454, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 52945200C122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 454
; LENGTH: 439892
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(439892)
; OTHER INFORMATION: n = A, T, C or G
US-10-087-192-454

Query Match 80.0%; Score 16.8; DB 13; Length 439892;
Best Local Similarity 90.0%; Pred. No. 3.9e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTGTGTGGGAGTTATTGAG 20
||||| |||||||

Db 86432 TTGTGTGGGAGTTATTGAG 86451

RESULT 12
US-10-424-599-85816/c
; Sequence 85816, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 85816
; LENGTH: 388
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_48505C.1
US-10-424-599-85816

Query Match 77.1%; Score 16.2; DB 13; Length 388;
Best Local Similarity 85.7%; Pred. No. 3.8e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TTGTGTGGGAGTTATTGAGT 21
||||| |||||||

Db 269 TTTTGTGGGAGGTATAGAGT 249

```
RESULT 13
US-10-027-632-185265
; Sequence 185265, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 185265
; LENGTH: 672
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-185265

Query Match          77.1%; Score 16.2; DB 13; Length 672;
Best Local Similarity 85.7%; Pred. No. 4e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 TTGTGTGGGGAGTTATTGAGT 21
Db      100 TAGTGGGGGAGTTATTGGGT 120

RESULT 14
US-10-027-632-185265
; Sequence 185265, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 185265
; LENGTH: 672
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-185265

Query Match          77.1%; Score 16.2; DB 13; Length 672;
Best Local Similarity 85.7%; Pred. No. 4e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 TTGTGTGGGGAGTTATTGAGT 21
Db      100 TAGTGGGGGAGTTATTGGGT 120
```

```
; ORGANISM: Human
US-10-027-632-185265

Query Match          77.1%; Score 16.2; DB 16; Length 672;
Best Local Similarity 85.7%; Pred. No. 4e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 TTGTGTGGGGAGTTATTGAGT 21
Db      100 TAGTGGGGGAGTTATTGGGT 120

RESULT 15
US-10-027-632-239897/c
; Sequence 239897, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 239897
; LENGTH: 703
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-239897

Query Match          77.1%; Score 16.2; DB 13; Length 703;
Best Local Similarity 85.7%; Pred. No. 4e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 TTGTGTGGGGAGTTATTGAGT 21
Db      248 TTGTGTGAGGATTGTGAGT 228

Search completed: April 24, 2004, 23:47:21
Job time : 184.296 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2004, 18:23:30 ; Search time 1634.63 Seconds
(without alignments)
383.638 Million cell updates/sec

Title: US-10-084-555a-115

Perfect score: 21

Sequence: 1 ttgtgtgggagttattgagt 21

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*

1: em_estbta.*

2: em_esthum.*

3: em_estin.*

4: em_estmu.*

5: em_estov.*

6: em_estopl.*

7: em_estro.*

8: em_htc.*

9: gb_est1.*

10: gb_est2.*

11: gb_htc.*

12: gb_est3.*

13: gb_est4.*

14: gb_est5.*

15: em_estfun.*

16: em_estom.*

17: em_gss_hum.*

18: em_gss_inv.*

19: em_gss_pln.*

20: em_gss_vrt.*

21: em_gss_fun.*

22: em_gss_man.*

23: em_gss_mus.*

24: em_gss_pro.*

25: em_gss_rod.*

26: em_gss_phg.*

27: em_gss_vri.*

28: gb_gss1.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18.4	87.6	306	14	N43966
2	18.4	87.6	498	28	BH899495
3	18.4	87.6	514	14	W03000
4	18.4	87.6	608	28	CC061126

5	18.4	87.6	770	13	BX106761
6	18.4	87.6	1416	29	AG076301
7	17.8	84.8	412	28	AQ278755
8	17.8	84.8	662	10	BE968245
9	17.8	84.8	728	29	AG170640
10	17.8	84.8	777	13	B0105673
11	17.4	82.9	305	13	B0495056
12	17.4	82.9	371	13	EU960021
13	17.4	82.9	613	13	B0761515
14	17.4	82.9	672	28	BZ155765
15	17.4	82.9	725	28	AZ200826
16	17.4	82.9	736	14	CF831623
17	17.4	82.9	803	9	AV755532
18	17.4	82.9	1063	28	CC193495
19	16.8	80.0	111	10	BG093013
20	16.8	80.0	178	28	AZ505013
21	16.8	80.0	180	10	BE165062
22	16.8	80.0	278	10	B8282403
23	16.8	80.0	315	12	BG467931
24	16.8	80.0	421	14	CD550566
25	16.8	80.0	478	12	BI419290
26	16.8	80.0	514	29	CE188017
27	16.8	80.0	530	28	AZ222364
28	16.8	80.0	539	28	AZ287267
29	16.8	80.0	548	14	CD545754
30	16.8	80.0	564	12	BJ382048
31	16.8	80.0	572	10	BF168334
32	16.8	80.0	586	28	BZ035221
33	16.8	80.0	598	12	EM601765
34	16.8	80.0	600	12	BG805255
35	16.8	80.0	600	12	B1986223
36	16.8	80.0	600	12	B1988907
37	16.8	80.0	601	12	EM625446
38	16.8	80.0	602	29	CE778003
39	16.8	80.0	606	12	BI417781
40	16.8	80.0	609	14	CF172352
41	16.8	80.0	618	28	AZ633952
42	16.8	80.0	643	14	CF914483
43	16.8	80.0	655	14	CF174956
44	16.8	80.0	668	29	CE565197
45	16.8	80.0	681	13	BU054731

ALIGNMENTS

RESULT 1
N43966
LOCUS
DEFINITION
YY28c01.r1 Soares melanocyte 2NbhM Homo sapiens cDNA Clone
IMAGE:272544 5', mRNA sequence.
N43966
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 306)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Narra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevas, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LLNL ; contact the

N43966 306 bp mRNA linear EST 07-FEB-1996

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: T7

High quality sequence stop: 260.

Location/Qualifiers

FEATURES

source

1. .306

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="GDB:3882186"

/db_xref="taxon:9606"

/clone="IMAGE:272544"

/sex="Male"

/tissue_type="melanocyte"

/lab_host="DH10B (ampicillin resistant)"

/clone_lib="Soares melanocyte 2NBHM"

/note="vector: p773D (Pharmacia) with a modified

polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA

was primed with a Not I - oligo(dT) primer [5',

TGTTACCAATCTGAAGCGGCGCGAGTTTCTTTTCTTTT 3'],

double-stranded cDNA was size selected, ligated to Eco RI

adapters (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of a modified p773 vector

(Pharmacia). Library constructed by Bento Soares and

M.Fatima Bonaldo. RNA from normal foreskin melanocytes

(FS374) was kindly provided by Dr. Anthony P. Albino."

ORIGIN

Query Match 87.6%; Score 18.4; DB 14; Length 306;

Best Local Similarity 95.0%; Pred. No. 1.1e+03;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TTGTGTGGGAGTTATTGAG 20

|||||

Db 123 TTGTGTGGGAGTTATTGAG 142

RESULT 2

BH899495/c

LOCUS

DEFINITION BH899495 498 bp DNA linear GSS 30-AUG-2002

tauri genomic clone oth08e07.g 5', genomic survey sequence.

ACCESSION BH899495

VERSION BH899495.1 GI:22550990

KEYWORDS GSS.

SOURCE Ostreococcus tauri

ORGANISM Ostreococcus tauri

REFERENCE 1 (bases 1 to 498)

AUTHORS Derelle, E., Ferraz, C., Lagoda, P., Eychenne, S., Cooke, R., Regad, F.,

Sabau, X., Courties, C., Delzeny, M., Demallie, J., Picard, A. and

Moreau, H.

TITLE DNA libraries for sequencing the genome of Ostreococcus tauri

(Chlorophytae, Prasinophyceae): the smallest free-living eukaryotic

cell

J. Phycol. 38 (6), 1150-1156 (2002)

CONTACT: Moreau H

LABORATORY: Arago

CNRS UMR 7628

BP 44, Avenue Fontale, 66651 Banyuls sur mer, France

Tel: (33) 468887309

Fax: (33) 468887398

Email: h.moreau@obs-banyuls.fr

Seq primer: reverse

Class: shotgun.

Location/Qualifiers

1. .498

/organism="Ostreococcus tauri"

/mol_type="genomic DNA"

/strain="OTH0595"

/db_xref="taxon:70448"

/clone="oth08e07.g"

/clone_lib="Ostreococcus tauri genomic shotgun library"

/note="vector: Bluescript; Site_1: EcoRV; Site_2: EcoRV;

FEATURES

source

Query Match 87.6%; Score 18.4; DB 14; Length 514;

Best Local Similarity 95.0%; Pred. No. 1.1e+03;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TTGTGTGGGAGTTATTGAG 20

ORIGIN

Query Match 87.6%; Score 18.4; DB 28; Length 498;

Best Local Similarity 95.0%; Pred. No. 1.1e+03;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TTGTGTGGGAGTTATTGAG 20

|||||

Db 411 TTGTGTGGGAGTTATTGAG 392

RESULT 3

W03000

LOCUS

DEFINITION

W03000

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1995)

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LiNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: mob.KEGA+ET

High quality sequence stop: 418.

Location/Qualifiers

1. .514

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="GDB:3884472"

/db_xref="taxon:9606"

/clone="IMAGE:291342"

/sex="Male"

/tissue_type="melanocyte"

/lab_host="DH10B (ampicillin resistant)"

/clone_lib="Soares melanocyte 2NBHM"

/note="vector: p773D (Pharmacia) with a modified

polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA

was primed with a Not I - oligo(dT) primer [5',

TGTTACCAATCTGAAGCGGCGCGAGTTTCTTTTCTTTT 3'],

double-stranded cDNA was size selected, ligated to Eco RI

adapters (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of a modified p773 vector

(Pharmacia). Library constructed by Bento Soares and

M.Fatima Bonaldo. RNA from normal foreskin melanocytes

(FS374) was kindly provided by Dr. Anthony P. Albino."

Shotgun library prepared after sonication of the genomic DNA. Blunt ligation in EcoRV site of Bluescript. Size selection of the inserts after agarose electrophoresis between 1 and 3 Kb."


```

Db      124 TTGTGTGGGGTGTATTGAG 143

RESULT 4
CC061126/c
LOCUS
DEFINITION MUGO_CH252P002Q3T7_H55_CD309_054 CHORI-252 Vervet Monkey Library
Cercopithecus aethiops genomic clone CH252-2U13, genomic survey
sequence.
ACCESSION CC061126
VERSION CC061126.1 GI:29790389
KEYWORDS GSS.
SOURCE Cercopithecus aethiops (African green monkey)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecinae; Cercopithecus.
REFERENCE 1 (bases 1 to 608)
AUTHORS Minhas,R., Zhang,X., Dore,C., Villeneuve,A., Lepage,P.,
Forgetta,V., McKee,K., Ophoff,R.A., Fairbanks,L.A., Freimer,N.B.,
Ervin,F.R., Palmour,R.M., Hudson,T.J. and Dewar, K.
TITLE UCLA/MUGO/St-Kitts Vervet Monkey Mapping Project
JOURNAL Unpublished (2003)
COMMENT Other GSSs: MUGO_CH252P002Q3SP6_H55_CD308_054
Contact: Dewar K
McGill University and Genome Quebec Innovation Centre
740 Dr. Penfield Room 7214, Montreal, QC, Canada, H3H 1A4
Tel: 514 398 3311 x00089
Fax: 514 398 1795
Email: ken.dewar@mcgill.ca
Plate: 2 row: J column: 13
Seq primer: T7 : TAATACGACTCACTATAGG
Class: BAC ends.

FEATURES
source
1..608
Location/Qualifiers
/organism="Cercopithecus aethiops"
/mol_type="genomic DNA"
/db_xref="taxon:9534"
/clone="CH252-2U13"
/sex="male"
/cell_type="White blood cell"
/dev_stage="Adult"
/clone_lib="CHORI-252 Vervet Monkey Library"
/note="Vector: pPARAC2.1; Site_1: EcoRI; Site_2: EcoRI;
Constructed by Michael Nefedov in Pieter de Jong's
laboratory at BACPAC Resources, Children's Hospital in
Oakland Research Institute."

ORIGIN
Query Match 87.6%; Score 18.4; DB 28; Length 608;
Best Local Similarity 95.0%; Pred. No. 1.1e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TTGTGTGGGGAGTTATTGAG 20
|||||
Db 196 TTGTGTGGGGTGTATTGAG 177

RESULT 5
BX106761
LOCUS
DEFINITION BX106761 Soares melanocyte 2NBHM Homo sapiens cDNA clone
IMAGE998B01600 ; IMAGE:272544, mRNA sequence.
ACCESSION BX106761
VERSION BX106761.1 GI:27834278
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 770)

AUTHORS Ebert,L., Heil,O., Hennig,S., Neubert,P., Partech,E., Peters,M.,
Radelof,U., Schneider,D. and Korn,B.
TITLE Human Unigeneset - RZPD3
JOURNAL Unpublished (2003)
COMMENT Contact: Ina Rolfs
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
RZPD; IMAGE998B01600.
RZPDLIB; I.M.A.G.E. cDNA Clone Collection;
Human Unigeneset - RZPD3 (RZPDLIB No.972)
http://www.rzpd.de/cloneCards/cgi-
bin/showLib.pl.cgi?response?libNo=972 Contact: Ina Rolfs
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Heubnerweg 6, D-14059 Berlin, Germany
Tel: +49 30 32639 101
Fax: +49 30 32639 111
www.rzpd.de
This clone is available royalty-free from RZPD;
contact RZPD (clone@rzpd.de) for further information. Seq primer:
M13r, Primer sequence: TTTCACACAGAAACAGCTATGAC.

FEATURES
source
1..770
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE998B01600 ; IMAGE:272544"
/sex="Male"
/tissue_type="melanocyte"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares melanocyte 2NBHM"
/note="Vector: pPVT3D (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAAGTGGGCGCGCGAGTTTATTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pPVT3 vector
(Pharmacia). Library constructed by Bento Soares and
M.Fatima Bonaldo. RNA from normal foreskin melanocytes
(FS374) was kindly provided by Dr. Anthony P. Albino."

ORIGIN
Query Match 87.6%; Score 18.4; DB 13; Length 770;
Best Local Similarity 95.0%; Pred. No. 1.1e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TTGTGTGGGGAGTTATTGAG 20
|||||
Db 124 TTGTGTGGGGTGTATTGAG 143

RESULT 6
AG076301
LOCUS
DEFINITION Pan troglodytes DNA, clone: PTB-070H11.F, genomic survey sequence.
ACCESSION AG076301
VERSION AG076301.1 GI:16628103
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of Library PTB
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 1416)
REFERENCE Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);

```

1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimbegsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.

PRIMERS

Sequencing: -21M13

LIBRARY

Vector : pKS145

R.Site 1 : SacI

R.Site 2 : SacI

Location/Qualifiers

1. 1416

/organism="Pan troglodytes"

/mol_type="genomic DNA"

/db_xref="taxon:9598"

/clone="PTB-070H11.F"

/sex="male"

/cell_type="lymphoblast"

/clone_lib="PTB Chimpanzee Male BAC Library"

ORIGIN

Query Match 87.6%; Score 18.4; DB 29; Length 1416;
Best Local Similarity 95.0%; Pred.No.1.1e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TGTGTGGGGAGTTATTGAGT 21

|||||

782 TGTGTGGGGAGTTATTGAGT 801

RESULT 7

AQ278755

LOCUS

DEFINITION AQ278755 412 bp DNA linear GSS 22-NOV-1998
CITBI-E1-2516F3.TR CITBI-E1 Homo sapiens genomic clone 2516F3,

genomic survey sequence.

ACCESSION AQ278755

VERSION AQ278755.1 GI:3904723

KEYWORDS GSS.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 412)

Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,

Berry,K., Granger,D., Suh,E., Wibler,C., Shizuya,H., Simon,M. and

Venter,J.C.

Use of a random human BAC End Sequence Database for Sequence-Ready

Map Building

Unpublished (1998)

Contact: Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: mdadams@tigr.org

Clones are available from Research Genetics (info@resgen.com). BAC

end search page:

http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html.

Seq primer: M13 Reverse

Class: BAC ends.

FEATURES

Location/Qualifiers

1. 412

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/clone="2516F3"

/sex="male"

/cell_type="sperm"

/clone_lib="CITBI-E1"

/note="Vector: pBelobAC11; Site_1: EcoRI; Site_2: EcoRI;

ORIGIN

Query Match 84.8%; Score 17.8; DB 28; Length 412;
Best Local Similarity 90.5%; Pred.No.2e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTGTGTGGGGAGTTATTGAGT 21

|||||

352 TTGTGTGGGGAGTTATTGAGT 372

RESULT 8

BE968245/c

LOCUS

DEFINITION BE968245 662 bp mRNA linear EST 04-OCT-2000
601648609R2 NIH_MGC_73 Homo sapiens cDNA clone IMAGE:3932426 3',
mRNA sequence.

ACCESSION BE968245

VERSION BE968245.1 GI:10578950

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 662)

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: gspbs-remail.nih.gov

Tissue Procurement: CLONTECH Laboratories, Inc.

cDNA Library Preparation: CLONTECH Laboratories, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLN)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCM770 row: a column: 03.

FEATURES

Location/Qualifiers

1. 662

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:3932426"

/lab_host="DH10B (T1 phage-resistant)"

/clone_lib="NIH_MGC_73"

/note="Organ: brain; Vector: pDNR-LIB (Clontech); Site_1:

SfiI (ggccgctcgcc); Site_2: SfiI (ggccattggcc); 5' and

3' adaptors were used in cloning as follows: 5' adaptor

sequence: 5'-CACGCCATTATGCC-3' and 3' adaptor sequence:

5'-ATTCTAGAGCGCGCGGCGGACATG-dt(30)BN-3' (where B = A,

C, or G and N = A, C, G, or T). Average insert size

1.35 kb (range 0.5-4.0 kb). 15/15 colonies contained

inserts by PCR. This library was enriched for full-length

clones and was constructed by Clontech Laboratories (Palo

Alto, CA)."

ORIGIN

Query Match 84.8%; Score 17.8; DB 10; Length 662;
Best Local Similarity 90.5%; Pred.No.2e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTGTGTGGGGAGTTATTGAGT 21

|||||

104 TTCTGTGGGGAGTTATTGAGT 84

RESULT 9

AG170640

LOCUS

DEFINITION AG170640 728 bp DNA linear GSS 09-JAN-2002
Pan troglodytes DNA, clone: RP43-039X20.TJ, genomic survey

sequence.

ACCESSION AG170640

```

VERSION      AGI70640.1  GI:16700318
KEYWORDS     GSS.
SOURCE       Pan troglodytes (chimpanzee)
ORGANISM     Pan troglodytes
             Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE    1
AUTHORS      Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
             Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE        BAC end sequences of library RPCI-43
JOURNAL      Unpublished
AUTHORS      2 (bases 1 to 728)
REFERENCE    Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
             Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE        Direct Submission
JOURNAL      Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
             and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
             1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
             (E-mail: chimpanzee@sc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/).
             Tel: 81-45-503-9111, Fax: 81-45-503-9170)
COMMENT      Clones are derived from the chimpanzee BAC library RPCI-43. This BAC
             end was generated during the R&D process and may have higher chance
             of clone tracking errors.
PRIMERS
Sequencing: TJ
LIBRARY      Vector      : pBACe3.6
             R.Site 1   : EcoRI
             R.Site 2   : EcoRI
             Location/Qualifiers
             1..728
             /organism="Pan troglodytes"
             /mol_type="genomic DNA"
             /db_xref="taxon:9598"
             /clone="RP43-039K20.TJ"
             /sex="male"
             /cell_type="lymphocytes"
             /clone_lib="RPCI-43 Chimpanzee Male BAC Library"

ORIGIN
Query Match      84.8%; Score 17.8; DB 29; Length 728;
Best Local Similarity 90.5%; Pred. No. 2e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTGTGGGGGAGTATTGAGT 21
Db 339 TTGTGGGGGAGTATTGAGT 359

RESULT 10
BU105673/c
LOCUS          777 bp  mRNA  linear  EST 25-NOV-2002
DEFINITION    60306477F1 CSQCCHL01 Gallus gallus cdna clone CHEST24j19 5', mRNA
sequence.
VERSION       BU105673.1  GI:25307627
KEYWORDS      EST.
SOURCE        Gallus gallus (chicken)
ORGANISM      Gallus gallus
             Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
             Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
             Phasianinae; Gallus.
REFERENCE     1 (bases 1 to 777)
AUTHORS      Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
             Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
TITLE        A Comprehensive Collection of Chicken cDNAs
JOURNAL      Curr Biol. 12 (22), 1965-1969 (2002)
MEDLINE      22335534
PUBMED       12445392
COMMENT      Contact: Simon Hubbard
             Department of Biomolecular Sciences
             University of Manchester Institute of Science and Technology
             (UMIST)

```

```

PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
Location/Qualifiers
1..777
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, Hisex"
/db_xref="taxon:9031"
/clone="CHEST24j19"
/tissue_type="whole embryo"
/dev_stage="20-21"
/lab_host="DH10B"
/clone_lib="CSQCCHL01"
note="Organ: whole embryo; Vector: pBluescript II KS(+);
Site 1: EcoRI; Site 2: NotI; Modification of pBluescript
II KS(+) [Stratagene] vector to accommodate cDNA produced
with the T-trimmed protocol (Construction of
uni-directionally cloned cDNA libraries from messenger RNA
for improved 3' end DNA sequencing by Glenn Fu, et al.
U.S. Patent # 6,287,624). Cut pBluescript II KS(+) with
NotI and EcoRI. Ligate in double stranded adaptor
containing BspI and BamHI sites
[5'ggcgcgtgcagccgcggtccgaaagaaag]
[5'aattcttttttcggtccggtgcgc]"

ORIGIN
Query Match      84.8%; Score 17.8; DB 13; Length 777;
Best Local Similarity 90.5%; Pred. No. 2e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTGTGGGGGAGTATTGAGT 21
Db 711 TTGTGGGGGAGTATTGAGT 691

RESULT 11
BQ495056
LOCUS          305 bp  mRNA  linear  EST 31-OCT-2002
DEFINITION    BQ495056 Pb0001 Paracoccidioides brasiliensis cDNA, mRNA sequence.
ACCESSION     BQ495056
VERSION       BQ495056.1  GI:24444420
KEYWORDS      EST.
SOURCE        Paracoccidioides brasiliensis
ORGANISM      Paracoccidioides brasiliensis
             Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
             Onygenales; mitosporic Onygenales; Paracoccidioides.
REFERENCE     1 (bases 1 to 305)
AUTHORS      Goldman,G.H., dos Reis Marques,E., Duarte Ribeiro,D.C., de Souza
             Bernardes,L.A., Quiapin,A.C., Vitorelli,P.M., Savoldi,M.,
             Semighini,C.P., de Oliveira,R.C., Nunes,L.R., Travaessos,L.R.,
             Puccia,R., Batista,W.L., Ferreira,L.E., Moreira,J.C.,
             Bogossian,A.P., Tekaia,F., Nobrega,M.P., Nobrega,F.G. and
             Goldman,M.H.
TITLE        Expressed sequence tag analysis of the human pathogen
             Paracoccidioides brasiliensis yeast phase: identification of
             putative homologues of Candida albicans virulence and pathogenicity
             genes
JOURNAL      Eukaryot. Cell 2 (1), 34-48 (2003)
COMMENT      Contact: Gustavo Henrique Goldman
             Laboratory of Molecular Biology
             Universidade de Sao Paulo - USP - FCFRP
             Av do Cafe S/N, CEP: 14040-903, Ribeirao Preto - SP, Brazil
             Email: ggoldman@usp.br
             Location/Qualifiers
             1..305
             /organism="Paracoccidioides brasiliensis"
             /mol_type="mRNA"
             /db_xref="taxon:121759"
             /clone_lib="Pb0001"

ORIGIN

```

```

Query Match      82.9%; Score 17.4; DB 13; Length 305;
Best Local Similarity 94.7%; Pred. No. 2.9e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TGTGTGGGGAGTATTGAG 20
DB 159 TGTGTGGGGAGTATTGAG 177

RESULT 12
LOCUS BU096021
DEFINITION tca-506 tca Trypanosoma carassii cDNA clone 03f14 5', mRNA
ACCESSION BU096021
VERSION BU096021.1 GI:25123745
KEYWORDS EST.
SOURCE Trypanosoma carassii
ORGANISM Trypanosoma carassii
          Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
          Trypanosoma.
REFERENCE 1 (bases 1 to 371)
AUTHORS Aguerro,F., Campo,V., Cremona,L., Jager,A., Di Noia,J.M.,
Ovarath,P., Sanchez,D.O. and Frasch,A.C.
TITLE Gene discovery in the freshwater fish parasite Trypanosoma
carassii: identification of trans-sialidase-like and mucin-like
genes
JOURNAL Infect. Immun. 70 (12), 7140-7144 (2002)
COMMENT Contact: Sanchez DO
Genomics and Bioinformatics
Instituto de Investigaciones Biotecnologicas
Av. Gral Paz S/N, INTI, Edificio 24, B 1650 KNA, San Martin, Buenos
Aires, Argentina
Tel: (54-11) 4580/7255/7
Fax: (54-11) 4752-9639
Email: dsanchez@lib.unsam.edu.ar
Sequences were basecalled with phred and vector was masked with
crossmatch (see http://www.phrap.org). Sequences were then trimmed
from both ends to remove low quality bases and masked vector.
Plate: 03 row: f column: 14
Seq primer: T7.
          Location/Qualifiers
          1..371
             /organism="Trypanosoma carassii"
             /mol_type="mRNA"
             /db_xref="taxon:38249"
             /clone="03f14"
             /dev_stage="blood trypanastigote"
             /lab_host="Goldfish (Carassius auratus)"
             /clone_lib="tca"
             /note="Vector: pSPort1; Blood trypanastigotes were
obtained from goldfish and cultured as described (Overath
et al. Parasitol Res (1998) 84:343) before obtaining total
RNA using Trizol. cDNA library construction was made from
polyA+ mRNA using a poly-dT oligonucleotide as primer. The
cDNAs were cloned in a oriented manner using a commercial
kit (SuperScript Plasmid System for cDNA Synthesis and
Plasmid Cloning, Life Technologies)."

ORIGIN
Query Match      82.9%; Score 17.4; DB 13; Length 371;
Best Local Similarity 94.7%; Pred. No. 2.9e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TGTGTGGGGAGTATTGAG 20
DB 200 TGTGTGGGGAGTATTGAG 218

RESULT 13
LOCUS BU761515
DEFINITION sas72905.y1 Gm-cl036 Glycine max cDNA clone SOYBEAN CLONE ID:

Query Match      82.9%; Score 17.4; DB 13; Length 613;
Best Local Similarity 94.7%; Pred. No. 2.9e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTGTGTGGGGAGTATTGA 19
DB 546 TTGTGTGGGGAGTATTGA 564

RESULT 14
LOCUS BZ155765
DEFINITION CH230-396G14.TJ CHORI-230 Segment 2 Rattus norvegicus genomic clone
CH230-396G14, genomic survey sequence.

Gm-cl036-9898 5' similar to TR:Q9SUG8 Q9SUG8 HYPOTHETICAL 64.6 KD
PROTEIN.; mRNA sequence.
BU761515
BU761515.1 GI:23727006
EST.
Glycine max (soybean)
Glycine max
Glycine
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
1 (bases 1 to 613)
Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Corryell,V.,
Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
Bowers,F., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
Schurk,R., Ritter,E., Kohn,T., Shin,T., Jackson,Y., Cardenas,M.,
McCann,R., Waterston,R. and Wilson,R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
This clone is available through: ResGen, Invitrogen Corp. 2130
South Memorial Parkway Huntsville, AL 35801 For further information
call: (800)-533-4363 or contact: ccu@resgen.com web site:
www.resgen.com
Seq primer: -40RP from Gibco
High quality sequence stop: 450.
          Location/Qualifiers
          1..613
             /organism="Glycine max"
             /mol_type="mRNA"
             /db_xref="taxon:3847"
             /clone="SOYBEAN CLONE ID: Gm-cl036-9898"
             /tissue_type="somatic embryos cultured on MSD 20"
             /lab_host="DH10B"
             /clone_lib="Gm-cl036"
             /note="Vector: pSPort1; Site 1: NotI; Site 2: SalI; This
cDNA library was constructed from mRNA isolated from
somatic embryos (age ranging from 2 months to 9 months)
cultured on MSD 20. The library was prepared using the
Life Technologies pSuperScript cDNA library construction
kit. Complementary DNA was synthesized from mRNA using a
poly (dT) sequence with a NotI restrictions site. SalI
linkers adapters were ligated to the blunt-ended cDNA
fragments followed by NotI digestion. The cDNA fragments
were directionally cloned into the NotI-SalI restriction
site of the pSPort1 vector. The ligated cDNA fragments
were transformed into E.coli ElectroMax DH10B host cells.
This library was constructed in the laboratory of Dr. Lila
Vodkin by Anu Khanna at the University of Illinois at
Urbana-Champaign. e-mail: l-vodkin@uiuc.edu"

```

```

ACCESSION   BZ155765
VERSION     BZ155765.1  GI:23796718
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1. (bases 1 to 672)
AUTHORS    Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K.,
            Shvartsbeyn,A., Gebregorgis,E., Overton,L., Russell,D., Chen,D.,
            Riggs,F., de Jong,P. and Fraser,C.M.
TITLE       Rat BAC End Sequences from Library CHORI-230 MboI segment
JOURNAL     Unpublished (1999)
COMMENT     Other GSSs: CH230-396G14.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.choi.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/oreringinformation.htm). BAC end
            page: http://www.tigr.org/tcb/bac\_ends/rat/bac\_end\_intro.html
            Plate: 396 row: G column: 14
            Seq primer: SP6
            Class: BAC ends.
FEATURES    source
            Location/Qualifiers
                1..672
                /organism="Rattus norvegicus"
                /mol_type="genomic DNA"
                /strain="BN/SNHsd/MCW"
                /db_xref="taxon:10116"
                /clones="CH230-396G14"
                /sex="Female"
                /cell_type="Brain"
                /clone_lib="CHORI-230 Segment 2"
                /note="Vector: pTARBAC1.3; Site 1: MboI; Site 2: MboI;
                CHORI-230 Rat (BN/SNHsd/MCW) BAC library produced by
                Pieter de Jong"
ORIGIN
Query Match      82.9%; Score 17.4; DB 28; Length 672;
Best Local Similarity 94.7%; Pred. NO. 2.9e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      3 GTGTGGGGAGTTATTGAGT 21
        |||||
Db      338 GTGTGGGGAGTTATTGAAT 356

RESULT 15
LOCUS    AZ200826
DEFINITION
SP_1011_A2_D11_SP6E Strongylocentrotus purpuratus, purple sea
urchin, sperm genomic BAC library Strongylocentrotus purpuratus
genomic clone Plate=1011 Col=22 Row=G, genomic survey sequence.
ACCESSION   AZ200826
VERSION     AZ200826.1  GI:8395740
KEYWORDS    GSS.
SOURCE      Strongylocentrotus purpuratus
ORGANISM    Strongylocentrotus purpuratus
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinida;
            Strongylocentrotidae; Strongylocentrotus.
            1 (bases 1 to 725)
REFERENCE   1. (bases 1 to 725)
AUTHORS    Cameron,R.A., Mahairas,G., Raat,J.P., Martinez,P., Biondi,T.R.,
            Swartzell,S., Wallace,J.C., Rouska,A.J., Livingston,B.T.,
            Wray,G.A., Ettensohn,C.A., Leinbach,H., Britten,R.J., Davidson,E.H.
and Hood,L.
            A sea urchin genome project: Sequence scan, virtual map, and
            additional resources
            Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)
            20402566
            MEDLINE
            PUBMED
            10920195
COMMENT     Contact: Cameron, RA, Davidson, EH, Hood, L
            Division of Biology 156-29
            California Institute of Technology
            Pasadena California 91125, USA
            Tel: (626) 395-8421
            Fax: (626) 793-3047
            Email: acameron@caltech.edu
            Plate: 1011 row: G column: 22
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 725.
FEATURES    source
            Location/Qualifiers
                1..725
                /organism="Strongylocentrotus purpuratus"
                /mol_type="genomic DNA"
                /db_xref="taxon:7668"
                /clone_lib="Plate=1011 Col=22 Row=G"
                /clone_lib="Strongylocentrotus purpuratus, purple sea
                urchin, sperm genomic BAC library"
                /note="Organ: sperm; Vector: BACE3.6; BAC Clones in E-Coli
                DH10B"
ORIGIN
Query Match      82.9%; Score 17.4; DB 28; Length 725;
Best Local Similarity 94.7%; Pred. NO. 2.9e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      3 GTGTGGGGAGTTATTGAGT 21
        |||||
Db      47 GTGTGGGGAGTTTGTGACT 65

Search completed: April 24, 2004, 21:10:46
Job time : 1639.63 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2004, 17:02:49 ; Search time 680.914 Seconds
(without alignments)
1464.047 Million cell updates/sec

Title: US-10-084-555A-116
Perfect score: 23
Sequence: 1 caccttcacaaaaaatcaatc 23

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb ba.*

2: gb htg.*

3: gb in.*

4: gb om.*

5: gb ov.*

6: gb pat.*

7: gb ph.*

8: gb pl.*

9: gb pr.*

10: gb ro.*

11: gb sts.*

12: gb sy.*

13: gb un.*

14: gb vi.*

15: em ba.*

16: em fun.*

17: em hum.*

18: em in.*

19: em mu.*

20: em om.*

21: em or.*

22: em ov.*

23: em pat.*

24: em ph.*

25: em pl.*

26: em ro.*

27: em sts.*

28: em un.*

29: em vi.*

30: em htg_hum.*

31: em htg_inv.*

32: em htg_other.*

33: em htg_mus.*

34: em htg_pln.*

35: em htg_rod.*

36: em htg_mam.*

37: em htg_vrt.*

38: em sy.*

39: em htgo_hum.*

40: em htgo_mus.*

41: em htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	20.4	88.7	108642	9	AC004451	AC004451 Homo sapi
C 2	20.4	88.7	131170	2	AC146437	AC146437 Pan trogl
C 3	19.8	86.1	85660	9	AL357148	AL357148 Human DNA
C 4	19.8	86.1	95593	10	BX470092	BX470092 Mouse DNA
C 5	19.8	86.1	184952	10	AC122358	AC122358 Mus muscu
C 6	19.8	86.1	254336	10	AC098838	AC098838 Genomic s
C 7	19.8	86.1	343019	2	AL391494	AL391494 Homo sapi
C 8	19.4	84.3	4634	2	AC014313	AC014313 Drosophil
C 9	19.4	84.3	50571	10	BX571685	BX571685 Mouse DNA
C 10	19.4	84.3	100029	10	AE014182_3	Continuation (4 of
C 11	19.4	84.3	110000	10	AE014182_2	Continuation (3 of
C 12	19.4	84.3	128341	9	AC091675	AC091675 Homo sapi
C 13	19.4	84.3	138032	2	AC141033	AC141033 Rattus no
C 14	19.4	84.3	138943	8	AC133709	AC133709 Medicago
C 15	19.4	84.3	158857	2	AC144928	AC144928 Medicago
C 16	19.4	84.3	176195	3	AC012165	AC012165 Drosophil
C 17	19.4	84.3	201858	2	AC025712	AC025712 Homo sapi
C 18	19.4	84.3	207432	3	AE003513	AE003513 Drosophil
C 19	19.4	84.3	208027	10	AC126691	AC126691 Mus muscu
C 20	19.4	84.3	209548	10	AC121612	AC121612 Mus muscu
C 21	19.4	84.3	249089	2	AC096295	AC096295 Rattus no
C 22	19.4	84.3	255484	2	AC094715	AC094715 Rattus no
C 23	19.4	84.3	304517	2	AC096282	AC096282 Rattus no
C 24	19.4	84.3	345997	2	AC096290	AC096290 Rattus no
C 25	19.4	82.6	147795	2	AC122443	AC122443 Mus muscu
C 26	18.8	81.7	5794	10	AE037169	AE037169 Mus muscu
C 27	18.8	81.7	7318	9	D7857652	D78577 Homo sapien
C 28	18.8	81.7	19087	6	AX345695	AX345695 Sequence
C 29	18.8	81.7	20078	8	AC007142	AC007142 Arabidops
C 30	18.8	81.7	37957	8	U17009	U17009 Phytophthor
C 31	18.8	81.7	40662	2	HSN4444	Z82248 Human DNA s
C 32	18.8	81.7	56781	2	AC087629	AC087629 Homo sapi
C 33	18.8	81.7	63771	9	AL590425	AL590425 Human DNA
C 34	18.8	81.7	63797	10	AF071080	AF071080 Mus muscu
C 35	18.8	81.7	66480	2	AC091568	AC091568 Homo sapi
C 36	18.8	81.7	69709	3	CEY6B3B	AL032655 Caenorhab
C 37	18.8	81.7	73509	2	AL137862	AL137862 Homo sapi
C 38	18.8	81.7	80053	2	AC022617	AC022617 Homo sapi
C 39	18.8	81.7	92345	9	AC121154	AC121154 Homo sapi
C 40	18.8	81.7	93220	8	AC006264	AC006264 Arabidops
C 41	18.8	81.7	107914	5	EX005373	EX005373 Zebrafish
C 42	18.8	81.7	110000	2	AC095248_3	Continuation (4 of
C 43	18.8	81.7	124669	2	AP004325	AP004325 Oryza sat
C 44	18.8	81.7	126595	9	AL160399	AL160399 Human DNA
C 45	18.8	81.7	128463	2	AP003619	AP003619 Oryza sat

ALIGNMENTS

RESULT 1
AC004451/c
LOCUS AC004451 108642 bp DNA linear PRI 04-JUN-2002
DEFINITION Homo sapiens PAC clone RP4-789N1 from 7q21, complete sequence.
ACCESSION AC004451
VERSION AC004451.2 GI:21322198
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 108642)
AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)

```

MEDLINE      99033792
PUBMED      9847074
REFERENCE    2 (bases 1 to 108642)
AUTHORS      Woessner,J., Mink,P., Hinds,K. and Stromwatt,C.
TITLE        The sequence of Homo sapiens PAC clone RP4-789N1
JOURNAL      Unpublished (2001)
REFERENCE    3 (bases 1 to 108642)
AUTHORS      Waterston,R.
TITLE        Direct Submission
JOURNAL      Submitted (20-MAR-1998) Department of Genetics,
              University, 4444 Forest Park Avenue, St. Louis,
              MO 63110
REFERENCE    4 (bases 1 to 108642)
AUTHORS      Waterston,R.
TITLE        Direct Submission
JOURNAL      Submitted (03-FEB-2000) Department of Genetics,
              University, 4444 Forest Park Avenue, St. Louis,
              MO 63110
REFERENCE    5 (bases 1 to 108642)
AUTHORS      Waterston,R.
TITLE        Direct Submission
JOURNAL      Submitted (04-JUN-2002) Department of Genetics,
              University, 4444 Forest Park Avenue, St. Louis,
              MO 63110
COMMENT      On Jun 4, 2002 this sequence version replaced gi
              108642.1 with gi 108642.2
              ----- Genome Center
              Center: Washington University Genome Sequencing
              Center
              Center code: WUGSC
              Web site: http://genome.wustl.edu/gsc
              Contact: sapiens@wustl.wustl.edu
              ----- Summary Statistics
              -----
              Center project name: H_DJ0789N01
              -----

```

	/db_xref="taxon:9606"	
	/chromosome="7"	
	/map="7q21"	
	/clone="RP4-789N1"	
	/clone_lib="RPCI-4"	
	1366..1394	
repeat_region	/rpt_family="AT_rich"	
repeat_region	3626..3692	
repeat_region	/rpt_family=" (GA)n"	
repeat_region	4049..4087	
repeat_region	/rpt_family=" (CA)n"	
repeat_region	4975..5033	
repeat_region	/rpt_family=" (GA)n"	
repeat_region	5063..5123	
repeat_region	/rpt_family="purine-rich"	
repeat_region	5481..5519	
repeat_region	/rpt_family="AT_rich"	
repeat_region	6368..6676	
repeat_region	/rpt_family="Alu"	
repeat_region	6820..7138	
repeat_region	/rpt_family="Alu"	
repeat_region	10792..10896	
repeat_region	/rpt_family="L1"	
repeat_region	11925..12010	
repeat_region	/rpt_family="L2"	
repeat_region	12140..12164	
repeat_region	/rpt_family=" (TA)n"	
repeat_region	12165..12467	
repeat_region	/rpt_family="Alu"	
repeat_region	12793..12823	
repeat_region	/rpt_family="AT_rich"	
repeat_region	14958..15358	
repeat_region	/rpt_family="Alu"	
repeat_region	15334..15386	
repeat_region	/rpt_family="Alu"	
repeat_region	16602..16720	
repeat_region	/rpt_family=" (TA)n"	
repeat_region	16808..17098	
repeat_region	/rpt_family="Alu"	
repeat_region	17939..17978	
repeat_region	/rpt_family=" (CA)n"	
repeat_region	18063..18146	
repeat_region	/rpt_family="MIR"	
repeat_region	18419..18453	
repeat_region	/rpt_family="AT_rich"	
repeat_region	18802..19345	
repeat_region	/rpt_family="L2"	
repeat_region	19352..19382	
repeat_region	/rpt_family="POLY_A"	
misc_feature	19366..19727	
misc_feature	/note="match to EST AA731945 (NID:g2753896) nw64g09.sl"	
misc_feature	19367..19692	
misc_feature	/note="similar to EST AA644250 (NID:g2569468) af63b06.sl"	
misc_feature	19368..19727	
misc_feature	/note="similar to EST AA779841 (NID:g2839172) af45g01.sl"	
misc_feature	19378..19791	
misc_feature	/note="similar to EST A1122887 (NID:g3538653) qb01f04.sl"	
misc_feature	19577..19791	
misc_feature	/note="match to EST AA320720 (NID:g1973048) "	
misc_feature	19611..19791	
misc_feature	/note="similar to EST AA171850 (NID:gl751104) zo98f09.sl"	
repeat_region	19966..20271	
repeat_region	/rpt_family="Alu"	
repeat_region	21538..21564	
repeat_region	/rpt_family="AT_rich"	
repeat_region	21605..21913	
repeat_region	/rpt_family="Alu"	
repeat_region	21924..22222	
repeat_region	/rpt_family="Alu"	
repeat_region	22714..22884	
repeat_region	/rpt_family="MIR"	
repeat_region	23335..23636	
repeat_region	/rpt_family="Alu"	

FEATURES
SOURCE

```

repeat_region 23637. 23807
/rpt_family="Alu"
repeat_region 23813. 23835
/rpt_family="T(AAA)n"
repeat_region 23845. 24126
/rpt_family="Alu"
repeat_region 24156. 24398
/rpt_family="L2"
repeat_region 24717. 25041
/rpt_family="Alu"
repeat_region 26317. 26414
/rpt_family="Alu"
repeat_region 26433. 26627
/rpt_family="Alu"
repeat_region 26703. 26729
/rpt_family="AT-rich"
repeat_region 26894. 26947
/rpt_family="AT-rich"
repeat_region 27824. 27859
/rpt_family="AT-rich"
repeat_region 28102. 28368
/rpt_family="MER2_type"
repeat_region 28341. 28500
/rpt_family="MER2_type"
repeat_region 28500. 30640
/rpt_family="MER2_type"
repeat_region 31284. 31318
/rpt_family="AT-rich"
repeat_region 31447. 31584
/rpt_family="AT-rich"
repeat_region 31928. 32008
/rpt_family="MER1_type"
repeat_region 35757. 35850
/rpt_family="CATA)n"
repeat_region 36523. 36559
/rpt_family="(TA)n"
repeat_region 37257. 37640
/rpt_family="(CA)n"
repeat_region 38236. 38328
/rpt_family="Retroviral"
repeat_region 38501. 39044
/rpt_family="AT-rich"

Query Match 88.7%; Score 20.4; DB 9; Length 108642;
Best Local Similarity 95.5%; Pred.No.3.4e-02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CACCTTCACAAAAAATCAAT 22
Db 52267 CCCCTTCACAAAAAATCAAT 52246

RESULT 2
AC146437/c 131170 bp DNA linear HTG 13-AUG-2003
LOCUS Pan troglodytes chromosome UNK clone RP43-15J3, *** SEQUENCING IN
DEFINITION PROGRESS ***, 44 unordered pieces.
ACCESSION AC146437.1 GI:33621017
VERSION AC146437.1 HTGS PHASE1.
KEYWORDS HTG; HTGS PHASE1.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
1 (bases 1 to 131170)
Wilson.R.K.
AUTHORS The sequence of Pan troglodytes clone
TITLE Unpublished JOURNAL
REFERENCE 2 (bases 1 to 131170)
AUTHORS Wilson.R.K.
TITLE Direct Submission
JOURNAL Submitted (13-AUG-2003) Genetics, Genome Sequencing Center, 4444
Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT

```

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: C_PT015J03

* NOTE: This is a 'working draft' sequence. It currently
* consists of 44 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1216: contig of 1216 bp in length
1316: gap of unknown length
1317: contig of 1423 bp in length
2740: gap of unknown length
2840: contig of 1722 bp in length
4561: gap of unknown length
4562: gap of unknown length
5865: contig of 1204 bp in length
5866: gap of unknown length
5966: contig of 1395 bp in length
7361: gap of unknown length
7361: contig of 1225 bp in length
8785: gap of unknown length
8786: contig of 1450 bp in length
10236: gap of unknown length
10236: contig of 1229 bp in length
11564: gap of unknown length
11565: contig of 1936 bp in length
13601: gap of unknown length
13701: contig of 2450 bp in length
16151: gap of unknown length
16250: contig of 2448 bp in length
18251: gap of unknown length
18251: contig of 1618 bp in length
18799: gap of unknown length
20416: contig of 2032 bp in length
22548: gap of unknown length
22549: contig of 2102 bp in length
22649: gap of unknown length
24751: contig of 1506 bp in length
24751: gap of unknown length
26357: contig of 1993 bp in length
26357: gap of unknown length
28450: contig of 2059 bp in length
28450: gap of unknown length
30609: contig of 2015 bp in length
30709: gap of unknown length
32724: contig of 1835 bp in length
32724: gap of unknown length
34558: contig of 2584 bp in length
34559: gap of unknown length
37342: contig of 2791 bp in length
37343: gap of unknown length
40233: contig of 1682 bp in length
40234: gap of unknown length
42015: contig of 1427 bp in length
42016: gap of unknown length
43542: contig of 3088 bp in length
43543: gap of unknown length
46730: contig of 2261 bp in length
46731: gap of unknown length
49091: contig of 1890 bp in length
49092: gap of unknown length
51081: contig of 3020 bp in length
51082: gap of unknown length
54201: contig of 1731 bp in length
54202: gap of unknown length
56032: contig of 2648 bp in length
56033: gap of unknown length
58780: contig of 2648 bp in length
56133

```



```

* 58781 58880: gap of unknown length
* 58881 61139: contig of 2259 bp in length
* 61140 61239: gap of unknown length
* 61240 63991: contig of 2752 bp in length
* 63992 64091: gap of unknown length
* 64092 66584: contig of 2493 bp in length
* 66585 66588: gap of unknown length
* 66589 69595: contig of 2911 bp in length
* 69596 69696: gap of unknown length
* 69697 73278: contig of 3583 bp in length
* 73279 73379: gap of unknown length
* 73380 78362: contig of 4984 bp in length
* 78363 78462: gap of unknown length
* 78463 81999: contig of 3437 bp in length
* 82000 86160: contig of 4161 bp in length
* 86161 91868: contig of 5608 bp in length
* 91869 96851: contig of 4883 bp in length
* 96852 96951: gap of unknown length
* 96952 101436: contig of 4484 bp in length
* 101437 101535: gap of unknown length
* 101536 107956: contig of 6421 bp in length
* 107957 108056: gap of unknown length
* 108057 115461: contig of 7405 bp in length
* 115462 123932: contig of 8371 bp in length
* 123933 131170: gap of unknown length
* 124033 131170: contig of 7138 bp in length.

```

FEATURES

source

```

1. .131170
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/chromosome="UNK"
/clone="RP43-15J3"
1. .1216
/note="assembly_name:Contig18"
1317. .2739
/note="assembly_name:Contig20"
2840. .4561
/note="assembly_name:Contig22"
4662. .5865
/note="assembly_name:Contig26"
5966. .7360
/note="assembly_name:Contig31"
7461. .8685
/note="assembly_name:Contig33"
8786. .10235
/note="assembly_name:Contig34"
10336. .11564
/note="assembly_name:Contig35"
11665. .13600
/note="assembly_name:Contig36"
13701. .16150
/note="assembly_name:Contig37"
16251. .18698
/note="assembly_name:Contig38"
18799. .20416
/note="assembly_name:Contig39"
20517. .22548
/note="assembly_name:Contig40"
22649. .24750
/note="assembly_name:Contig42"
24851. .26356
/note="assembly_name:Contig43"
26457. .28449
/note="assembly_name:Contig44"
28550. .30608
/note="assembly_name:Contig45"
30709. .32723
/note="assembly_name:Contig46"

```

```

misc_feature 32824. .34658
/note="assembly_name:Contig47"
misc_feature 34759. .37342
/note="assembly_name:Contig48"
misc_feature 37443. .40233
/note="assembly_name:Contig49"
misc_feature 40334. .42015
/note="assembly_name:Contig50"
misc_feature 42116. .43542
/note="assembly_name:Contig51"
misc_feature 43643. .46730
/note="assembly_name:Contig52"
misc_feature 46831. .49091
/note="assembly_name:Contig53"
misc_feature 49192. .51081
/note="assembly_name:Contig54"
misc_feature 51182. .54201
/note="assembly_name:Contig55"
misc_feature 54302. .56032
/note="assembly_name:Contig56"
misc_feature 56133. .58780
/note="assembly_name:Contig57"
misc_feature 58881. .61139
/note="assembly_name:Contig58"
misc_feature 61240. .63991
/note="assembly_name:Contig59"
misc_feature 64092. .66584
/note="assembly_name:Contig60"
misc_feature 66685. .69595
/note="assembly_name:Contig61"
misc_feature 69696. .73278
/note="assembly_name:Contig62"
misc_feature 73379. .78362
/note="assembly_name:Contig63"

```

Query Match 88.7%; Score 20.4; DB 2; Length 131170;

Best Local Similarity 95.5%; Pctd. No. 3.2e+02;

Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CACCTTCACAAAAAATCAAT 22

Db 73402 CCGCTTCACAAAAAATCAAT 73381

RESULT 3

AL357148/c

LOCUS

DEFINITION

Human DNA sequence from clone Rp11-739D18 on chromosome 10,

complete sequence.

ACCESSION

AL357148

VERSION

AL357148.22

KEYWORDS

HTG.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 85660)

AUTHORS

Bird,C.

TITLE

Direct Submission

JOURNAL

Submitted (02-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

humquery@sanger.ac.uk

On May 2, 2003 this sequence version replaced gi:20338435.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:

Emi, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; information

on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr10>

RP11-739D18 is from the library RP11-11.3 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBAC3.6

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

FEATURES

Location/Qualifiers
1..85660
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-739D18"
/clone_lib="RP11-11.3"

ORIGIN

Query Match 86.1%; Score 19.8; DB 9; Length 85660;
Best Local Similarity 91.3%; Pred. No. 6e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CACCTTCACAAAAAATCAATC 23
|||||
Db 33007 CACCTTCACAAAAAATCAATC 32985

RESULT 4
BX470092/c 95593 bp DNA linear ROD 30-NOV-2003
LOCUS Mouse DNA sequence from clone RP23-230K20 on chromosome X, complete sequence.

ACCESSION BX470092
VERSION BX470092.12 GI:38568114
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 95593)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE Ellwood, M.
JOURNAL Direct Submission

Submitted (29-NOV-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Nov 30, 2003 this sequence version replaced gi:35209691.

Sequence from the Mouse Genome Sequencing Consortium whole genome shotgun may have been used to confirm this sequence. Sequence data from the whole genome shotgun alone has only been used where it has a phred quality of at least 30.

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-230K20 is from the RP11-23 Mouse BAC Library constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBAC3.6.

FEATURES

Location/Qualifiers
1..95593
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="X"
/clone="RP23-230K20"
/clone_lib="RP11-23"

ORIGIN

Query Match 86.1%; Score 19.8; DB 10; Length 95593;
Best Local Similarity 91.3%; Pred. No. 5.9e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CACCTTCACAAAAAATCAATC 23
|||||
Db 7769 CACTTTCACAAAAAATCAATC 7747

RESULT 5
AC122358

LOCUS Mus musculus BAC clone RP23-406B6 from chromosome 17, complete sequence.

ACCESSION AC122358
VERSION AC122358.2 GI:23334930
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 184952)
AUTHORS Ali, J. and Cotton, M.
TITLE The sequence of Mus musculus BAC clone RP23-406B6
JOURNAL Unpublished (2001)

REFERENCE 2 (bases 1 to 184952)
AUTHORS Wilson, R.
TITLE Sequencing of Mus musculus
JOURNAL Unpublished (2001)

REFERENCE 3 (bases 1 to 184952)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (23-MAY-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE 4 (bases 1 to 184952)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (27-SEP-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE 5 (bases 1 to 184952)
AUTHORS Wilson, R.
TITLE Direct Submission

repeat_region	/rpt_family="B4"	1164. .1447
repeat_region	/rpt_family="B4"	8177. .8303
repeat_region	/rpt_family="Alu"	8304. .8340
repeat_region	/rpt_family="B4"	8837. .8988
repeat_region	/rpt_family="B4"	11144. .11620
repeat_region	/rpt_family="MaLR"	12852. .13350
repeat_region	/rpt_family="L1"	13551. .13672
repeat_region	/rpt_family="ERVK"	13673. .13994
repeat_region	/rpt_family="L1"	14043. .14122
repeat_region	/rpt_family="L2"	14593. .14923
repeat_region	/rpt_family="L1"	14948. .15089
repeat_region	/rpt_family="B4"	
repeat_region	/rpt_family="B4"	38495. .38650
repeat_region	/rpt_family="MER1_type"	38672. .39018
repeat_region	/rpt_family="MaLR"	39740. .39798
repeat_region	/rpt_family="L1"	40005. .40277
repeat_region	/rpt_family="L1"	49107. .49224
repeat_region	/rpt_family="B2"	49411. .49560
repeat_region	/rpt_family="B2"	49903. .49947
repeat_region	/rpt_family="B4"	49948. .50566
repeat_region	/rpt_family="RMER13A"	50849. .50722
repeat_region	/rpt_family="B4"	50723. .51081
repeat_region	/rpt_family="L1"	51082. .51133

```

repeat_region      /rpt_family="B4"
51454..51740
/rpt_family="MaLR"
repeat_region      51951..52021
/rpt_family="MaLR"
repeat_region      52028..52418
/rpt_family="MaLR"
repeat_region      53094..53594
/rpt_family="MaLR"
repeat_region      53821..53915
/rpt_family="CR1"
repeat_region      54087..54153
/rpt_family="L2"
repeat_region      54626..54798
/rpt_family="L1"
repeat_region      55017..55125
/rpt_family="L1"
repeat_region      55768..55959
/rpt_family="B2"
repeat_region      55979..56066
/rpt_family="L1"
repeat_region      56077..56166
/rpt_family="B4"

Query Match      86.1%; Score 19.8; DB 10; Length 184952;
Best Local Similarity 91.3%; Pred. No. 5.1e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CACCTTCACAAAAAATCAATC 23
|||||
Db 154763 CACATTCACAAAAAATCAATC 154785
|||||

RESULT 6
AC098838/c
LOCUS      AC098838      254336 bp      DNA      linear      ROD 12-JUN-2002
DEFINITION Genomic sequence for Mus musculus, clone RP23-27N1, from chromosome
17, complete sequence.
ACCESSION      AC098838
VERSION      AC098838.3 GI:21392464
KEYWORDS      HTG.
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 254336)
McCombie,W.R., Spiegel,L., de la Baside,M., Preston,R.,
Ferraro,K., Kuit,K., Nascimento,L., Zucavern,T., Baija,V.,
Bell,M., Baker,J., Miller,B., Katzenberger,F., Muller,S., King,L.,
Sullivan,P., Yang,C., Dike,S., Palmer,L., O'Shaughnessy,A. and
Dedhia,N.
Genomic sequence for Mus musculus, clone RP23-27N1, from chromosome
17, complete sequence
Unpublished
2 (bases 1 to 254336)
McCombie,W.R.
Direct Submission
Submitted (03-NOV-2001) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
3 (bases 1 to 254336)
McCombie,W.R.
Direct Submission
Submitted (12-JUN-2002) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
On Jun 12, 2002 this sequence version replaced gi:16973731.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest.
Clone RP23-27N1 overlaps clone RP23-8L6 (AC104519) from base 231689
to base 254336. The overlap is from base 1 to base 22648 on
RP23-8L6.
FEATURES             Location/Qualifiers
     source            1..254336
                        /organism="Mus musculus"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:10090"
                        /chromosome="17"
                        /clone="RP23-27N1"
                        /clone_lib="RPCL-23"
                        complement(92247..92302)
     misc_feature      /note="We believe the assembly to be correct. The sequence
                        is a polynucleotide repeat (CCTT) in which the exact
                        number of repeats is unknown. One subclone in the region
                        shows one fewer CCTT repeat than that represented by the
                        assembly."
                        complement(113713..113724)
     misc_feature      /note="We believe the assembly to be correct. The
                        sequence is a mononucleotide (T) repeat in which the exact
                        number of Ts is unknown. One subclone in the region shows
                        one additional T compared to that represented by the
                        assembly."
                        148545..148615
     misc_feature      /note="We believe the assembly to be correct. The
                        sequence is covered solely by sequences amplified from the
                        BAC DNA template. The sequence is high quality."
                        157849..157938
     misc_feature      /note="We believe the assembly to be correct. The
                        sequence is covered solely by sequences amplified from the
                        BAC DNA template. The sequence is high quality."
                        158679..158945
     misc_feature      /note="We believe the assembly to be correct. The
                        sequence is a polynucleotide repeat which is covered by
                        sequences generated from the transposition of plasmids
                        spanning the region. The assembly is consistent with
                        restriction digest information."
                        221197..221248
     misc_feature      /note="We believe the assembly to be correct. The
                        sequence is a dinucleotide (TnGn) repeat in which the
                        exact number of repeat copies is unknown. The assembly is
                        consistent with restriction digest information."
                        229839..229879
     misc_feature      /note="We believe the assembly to be correct. The
                        sequence is covered solely by sequences which were
                        amplified from the BAC DNA template. The sequence is high
                        quality."

ORIGIN
Query Match      86.1%; Score 19.8; DB 10; Length 254336;
Best Local Similarity 91.3%; Pred. No. 4.8e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CACCTTCACAAAAAATCAATC 23
|||||
Db 172818 CACATTCACAAAAAATCAATC 172796
|||||

RESULT 7
AL391494
LOCUS      AL391494      343019 bp      DNA      linear      HTG 02-MAY-2003
DEFINITION Homo sapiens chromosome 10 clone RP13-263N18, *** SEQUENCING IN
PROGRESS ***, 12 unordered pieces.
ACCESSION      AL391494
VERSION      AL391494.9 GI:30348811
KEYWORDS      HTG; HTGS_PHASE1.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 343019)
Sims,S.

```


Db 2202 CATTCAACAAAAAATCAATC 2182

RESULT 9

BX571685
 LOCUS Mouse DNA sequence from clone RP23-29D24 on chromosome 4, complete sequence.
 DEFINITION
 ACCESSION BX571685.1 GI:32567440
 VERSION
 KEYWORDS HTG.
 SOURCE Mus musculus (house mouse)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 50571)
 AUTHORS North.P., Leaves.N., Greystrom.J., Coppola.M., Manjunath.S., Russell.E., Smith.M., Strachan.G., Tofts.C., Boal.E., Cobley.V., Hunter.G., Kimberley.C., Thomas.D., Cave-Berry.L., Weston.P. and Botcherby.M.R.M.
 TITLE Direct Submission
 JOURNAL Submitted (12-JUL-2003) Mouse Sequencing Group, HGMP-RC, Hinxton, Cambridge, CB10 1SB, UK. E-mail enquiries:- mtbotche@hmp.mrc.ac.uk or pnothe@hmp.mrc.ac.uk
 REMARK HGMP-RC part of the UK Mouse Sequencing Consortium
 COMMENT ----- Genome Center
 Center: UK Medical Research Council
 Center code: UK-MRC
 Web site: <http://mrcseq.har.mrc.ac.uk>
 Contact: mouse@har.mrc.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Swt., SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-29D24 is from the RPI-23 Mouse BAC Library

constructed by the group of Pieter de Jong.

For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACe3.6

Sequence from the Mouse Genome Sequencing Consortium whole genome shotgun may have been used to confirm this sequence. Sequence data from the whole genome shotgun alone has only been used where it has a phred quality of at least 30.

FEATURES

Source

1. 50571
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="4"
 /clone="RP23-29D24"
 /clone_lib="RPI-23"

ORIGIN

Query Match 84.3%; Score 19.4; DB 10; Length 50571;
 Best Local Similarity 95.2%; Pred. No. 9.5e+02;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CACCTTCACAAAAAATCAA 21
 Db 12491 CACCTTCACAAAAAATAAA 12511

RESULT 10

AE014182_3
 WPCOMMENT
 Sequence split into 4 fragments LOCUS AE014182 Accession AE014182
 Fragment Name Begin End
 AE014182_0 1 110000
 AE014182_1 100001 210000
 AE014182_2 200001 310000
 AE014182_3 300001 400029
 Continuation (4 of 4) of AE014182 from base 300001 (AE014182 Mus musculus piebald delet

Query Match 84.3%; Score 19.4; DB 10; Length 100029;
 Best Local Similarity 95.2%; Pred. No. 8.3e+02;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 CCTTCACAAAAAATCAATC 23
 Db 143 CCTTCACAAAAAATCAATC 163

RESULT 11

AE014182_2
 WPCOMMENT
 Sequence split into 4 fragments LOCUS AE014182 Accession AE014182
 Fragment Name Begin End
 AE014182_0 1 110000
 AE014182_1 100001 210000
 AE014182_2 200001 310000
 AE014182_3 300001 400029
 Continuation (3 of 4) of AE014182 from base 200001 (AE014182 Mus musculus piebald delet

Query Match 84.3%; Score 19.4; DB 10; Length 110000;
 Best Local Similarity 95.2%; Pred. No. 8.1e+02;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 CCTTCACAAAAAATCAATC 23
 Db 100143 CCTTCACAAAAAATCAATC 100163

RESULT 12

AC091675
 LOCUS Homo sapiens chromosome 8, clone RP11-384C12, complete sequence.
 DEFINITION
 AC091675
 ACCESSION AC091675.5 GI:20377014
 VERSION
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 128341)
 AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,B.
 TITLE Homo sapiens chromosome 8, clone RP11-384C12
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 128341)
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,B., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M., Collins,S., Collumore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K., Lamasares,R., Landers,T., Lehotzky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrum,J., Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,


```

unsure      13384..13392
            /note=<30 qual SNGL region"
unsure      13434..13439
            /note=<30 qual SNGL region"
unsure      13447..13451
            /note=<30 qual SNGL region"
unsure      13462..13520
            /note=<30 qual SNGL region"
            /rpt_family="MIR"
repeat_region 14284..14296
            /note="Single clone coverage"
            /rpt_family="AluX"
repeat_region 15624..15923
            /rpt_family="MIR"
repeat_region 16388..16599
            /rpt_family="MIR"
repeat_region 17122..17712
            /rpt_family="L16A"
repeat_region 18019..18126
            /rpt_family="MIR"
repeat_region 19362..19734
            /rpt_family="THE1C"
repeat_region 22231..22382
            /rpt_family="L16A"
repeat_region 22611..22806
            /rpt_family="L16C/D"
repeat_region 23677..23714
            /rpt_family="(TC)n"

Query Match      84.3%; Score 19.4; DB 9; Length 128341;
Best Local Similarity 95.2%; Pred. No. 7.9e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CACCTTCACAAAAAATCA 21
Db 57593 CATCTTCACAAAAAATCA 57613

RESULT 13
AC141033
LOCUS      AC141033      138032 bp      DNA      linear      HTG 27-MAR-2003
DEFINITION Rattus norvegicus clone CH230-469L1, *** SEQUENCING IN PROGRESS
ACCESSION  AC141033
VERSION     AC141033.1 GI:28875892
KEYWORDS    HTG, HTGS_PHASE1.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 138032)
            Murny,D., Maris., Metzker,M., Lee., Abramson,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavaros,I., Cesar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebreyogis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Rume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,

```

```

Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuwa,L., Louisedge,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M., McNeill,T., Meenen,E., Milosavljevic,A.,
Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K.,
Morris,S., Muidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D.,
Newton,N., Nguyen,N., Norris,S., Nwaokemelehen,C., Okwundu,G.,
Olanpungsoon,A., Fal,S., Parks,K., Pasternak,S., Paul,H.,
Perez,A., Perez,L., Pfankoch,C., Plopper,F., Poindexter,A.,
Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E.,
Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y.,
Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A.,
Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S.,
Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A.,
Sisson,I., Sitter,C.D., Smajs,D., Speed,A., Sodergren,E.,
Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A.,
Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S.,
Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villagana,D.,
Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J.,
Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wlaczky,R.,
Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S.,
Yen,J., Yoon,D., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X.,
Zhao,S., Dunn,D., von Niederhauser,A., Weiss,R., Smith,D.R.,
Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 138032)
Worley,K.C.
Direct Submission
Submitted (07-MAR-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 138032)
Worley,K.C.
Direct Submission
Submitted (27-MAR-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
----- Genome Center
Center: Baylor College of Medicine
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GXJF
Center clone name: CH230-469L1
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 114059 bases at least Q40
Consensus quality: 120132 bases at least Q30
Consensus quality: 124173 bases at least Q20
Estimated insert size: 117098; sum-of-contrigs estimation
Quality coverage: 2x in Q20 bases; sum-of-contrigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This is a 'working draft' sequence. It currently
* consists of 38 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1066: contig of 1066 bp in length
* 1067 1166: gap of unknown length
* 1167 2571: contig of 1405 bp in length
* 2572 2671: gap of unknown length

```


* 2672 3720: contig of 1049 bp in length
* 3721 3820: gap of unknown length
* 3821 5517: contig of 1697 bp in length
* 5518 5617: gap of unknown length
* 5618 6890: contig of 1273 bp in length
* 6891 6990: gap of unknown length
* 6991 8090: contig of 1100 bp in length
* 8091 8190: gap of unknown length
* 8191 9957: contig of 1767 bp in length
* 9958 10057: gap of unknown length
* 10058 11291: contig of 1234 bp in length
* 11292 11391: gap of unknown length
* 11392 13813: contig of 2422 bp in length
* 13814 15633: contig of 1720 bp in length
* 15634 15733: gap of unknown length
* 15734 17114: contig of 1381 bp in length
* 17115 17215: gap of unknown length
* 17216 18792: contig of 1578 bp in length
* 18793 18892: gap of unknown length
* 18893 20497: contig of 1605 bp in length
* 20498 20597: gap of unknown length
* 20598 22665: contig of 2068 bp in length
* 22666 22765: gap of unknown length
* 22766 25618: contig of 2853 bp in length
* 25619 25718: gap of unknown length
* 25719 28335: contig of 2617 bp in length
* 28336 28435: gap of unknown length
* 28436 30744: contig of 2309 bp in length
* 30745 30844: gap of unknown length
* 30845 33350: contig of 2506 bp in length
* 33351 35731: gap of unknown length
* 35732 36831: gap of unknown length
* 36832 39087: contig of 2256 bp in length
* 39088 39187: gap of unknown length
* 39188 42302: contig of 3115 bp in length
* 42303 42403: gap of unknown length
* 42404 45290: contig of 2888 bp in length
* 45291 45390: gap of unknown length
* 45391 47429: contig of 2039 bp in length
* 47430 47529: gap of unknown length
* 47530 50533: contig of 3004 bp in length
* 50534 50633: gap of unknown length
* 50634 54177: contig of 3544 bp in length
* 54178 54277: gap of unknown length
* 54278 58295: contig of 4018 bp in length
* 58296 58395: gap of unknown length
* 58396 62268: contig of 3873 bp in length
* 62269 62368: gap of unknown length
* 62369 66603: contig of 4235 bp in length
* 66604 67003: gap of unknown length
* 67004 71708: contig of 5005 bp in length
* 71709 71808: gap of unknown length
* 71809 77433: contig of 5625 bp in length
* 77434 77533: gap of unknown length
* 77534 83635: contig of 6102 bp in length
* 83636 83735: gap of unknown length
* 83736 86148: contig of 4413 bp in length
* 86149 88248: gap of unknown length
* 88249 92548: contig of 4200 bp in length
* 92549 92549: gap of unknown length
* 92550 98362: contig of 5814 bp in length
* 98363 98463: gap of unknown length
* 98464 106812: contig of 8350 bp in length
* 106813 106912: gap of unknown length
* 106913 117994: contig of 11082 bp in length
* 117995 118094: gap of unknown length
* 118095 128775: contig of 10681 bp in length
* 128776 128875: gap of unknown length
* 128876 138032: contig of 9157 bp in length.

Location/Qualifiers
1..138032

FEATURES
source

/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-469L1"

ORIGIN

Query Match 84.3%; Score 19.4; DB 2; Length 138032;
Best Local Similarity 95.2%; Pred. No. 7.7e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CACCTTCACAAAAAATCAA 21

|||||
Db 89154 CACCTTCACAAAAAATCAA 89174

RESULT 14

AC133709/c

LOCUS

Medicago truncatula clone mth2-7b3, complete sequence.

DEFINITION

AC133709

ACCESSION

AC133709.8

VERSION

KEYWORDS

SOURCE

ORGANISM

Medicago truncatula (barrel medic)

Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;

Medicago.

1 (bases 1 to 138943)

Shaul, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,

Cook, D., Kim, D. and Roe, B.A.

Medicago truncatula BAC Clone mth2-7b3

Unpublished

2 (bases 1 to 138943)

Shaul, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,

Cook, D., Kim, D. and Roe, B.A.

Direct Submission

Submitted (17-SEP-2002) Department of Chemistry And Biochemistry,

The University Of Oklahoma, 620 Farrington Oval, Room 208, Norman,

OK 73019, USA

3 (bases 1 to 138943)

Shaul, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,

Cook, D., Kim, D. and Roe, B.A.

Direct Submission

Submitted (26-FEB-2003) Department of Chemistry And Biochemistry,

The University Of Oklahoma, 620 Farrington Oval, Room 208, Norman,

OK 73019, USA

4 (bases 1 to 138943)

Shaul, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,

Cook, D., Kim, D. and Roe, B.A.

Direct Submission

Submitted (09-APR-2003) Department of Chemistry And Biochemistry,

The University Of Oklahoma, 620 Farrington Oval, Room 208, Norman,

OK 73019, USA

5 (bases 1 to 138943)

Shaul, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,

Cook, D., Kim, D. and Roe, B.A.

Direct Submission

Submitted (10-APR-2003) Department of Chemistry And Biochemistry,

The University Of Oklahoma, 620 Farrington Oval, Room 208, Norman,

OK 73019, USA

6 (bases 1 to 138943)

Shaul, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,

Cook, D., Kim, D. and Roe, B.A.

Direct Submission

Submitted (24-APR-2003) Department of Chemistry And Biochemistry,

The University Of Oklahoma, 620 Farrington Oval, Room 208, Norman,

OK 73019, USA

On Apr 9, 2003 this sequence version replaced gi:28273426.

----- Genome Center

Center: Department of Chemistry And Biochemistry

The University Of Oklahoma

Center code: UOKNOR

ORIGIN

Query Match 84.3%; Score 19.4; DB 2; Length 158857;
Best Local Similarity 95.2%; Pred. No. 7.5e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CACCTTCACAAAAAATCAA 21
|||||
Db 34270 CACCTTCACAAAAAATCAA 34250

Search completed: April 24, 2004, 19:25:26
Job time : 685.914 secs

CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention

XX SQ Sequence 19087 BP; 6238 A; 296 C; 3614 G; 8939 T; 0 U; 0 Other;

Query Match 81.7%; Score 18.8; DB 6; Length 19087;
 Best Local Similarity 90.9%; Pred. No. 5.8e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ACCCTTCACAAAAAATCAATC 23

DB 5763 ACCCTTCACAAAAAATTCATC 5742

RESULT 2

ABL33599/C
 ID ABL33599 standard; DNA; 5487 BP.

XX ABL33599;

DT 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 1572.

XX Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianemic; cytosine; neotropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
 KW ds.

XX Homo sapiens.

OS WO200200928-A2.

PN 03-JAN-2002.

PD 02-JUL-2001; 2001WO-EP007537.

PF 30-JUN-2000; 2000DE-01032529.

PR 01-SEP-2000; 2000DE-01043826.

XX (EPIG-) EPIGENOMICS AG.

PA Olek A, Piepenbrock C, Berlin K;

PI WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful for
 PT diagnosis and treatment of diseases associated with abnormal cytosine
 PT methylation.

XX Claim 1; SEQ ID NO 1572; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention

XX SQ Sequence 5487 BP; 1608 A; 133 C; 1240 G; 2506 T; 0 U; 0 Other;

Query Match 80.0%; Score 18.4; DB 6; Length 5487;
 Best Local Similarity 95.0%; Pred. No. 7.8e+02;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ACCCTTCACAAAAAATCAAA 21

DB 5375 ACCCTTCATAAAAAAATCAAA 5356

RESULT 3

ABN80072/C

ID ABN80072 standard; DNA; 5771 BP.

XX AC ABN80072;

DT 15-JUL-2002 (first entry)

DE Human chemically modified disease associated gene SEQ ID NO 89.

XX Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;
 KW heart disease; epilepsy; histone deacetylation; muscular dystrophy;
 KW dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;
 KW antidiabetic; cytosine; anticonvulsant; ds.

XX Homo sapiens.

OS Synthetic.

XX WO200200927-A2.

PN 03-JAN-2002.

PD 02-JUL-2001; 2001WO-EP007536.

PF 30-JUN-2000; 2000DE-01032529.

PR 01-SEP-2000; 2000DE-01043826.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130908/17.

XX Novel nucleic acid useful for diagnosis and therapy of diseases
 PT associated with development genes such as diabetes, comprises a sequence
 PT of a segment of chemically pretreated DNA of genes associated with
 PT development.

XX Claim 1; SEQ ID NO 89; 27pp; English.

XX The invention relates to a nucleic acid (I) comprising a sequence at
 CC least 18 bases in length of a segment of chemically pretreated DNA (II)
 CC of genes associated with development selected from 87 genes listed in the
 CC specification such as ACCPN, ADFN, or AFD1 and comprising one of 350
 CC sequences (ABN7984-ABN8033) or their complements. The invention is
 CC useful for the diagnosis or therapy of diseases associated with
 CC development genes, in particular disease related to homeobox containing
 CC genes (HOX), like diabetes, cancer, apoptosis related diseases, syndromes
 CC associated with congenital heart disease, epilepsy, diseases related to
 CC histone deacetylation, Currarino syndrome, diseases related with the
 CC development of the brain and limb girdle muscular dystrophy and dwarfism.
 CC Oligomers specific to each of the genes are useful for detecting the
 CC methylation state of all CpG dinucleotides within the 350 sequences or
 CC (II) and their complementary sequences, as primer oligonucleotides for
 CC the amplification of the 350 sequences, (II) and/or their complements and
 CC as oligomer probes for detecting the cytosine methylation state and/or
 CC single nucleotide polymorphisms (SNPs). Note: The sequence data for this
 CC patent did not form part of the printed specification but is based on
 CC sequence information supplied to Derwent by the European Patent Office

XX SQ Sequence 5771 BP; 1877 A; 37 C; 980 G; 2877 T; 0 U; 0 Other;

Query Match 80.0%; Score 18.4; DB 6; Length 5771;
 Best Local Similarity 95.0%; Pred. No. 7.8e+02;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ACCCTTCACAAAAAATCAAA 21

Human immune system associated gene SEQ ID NO: 547.

Human; immune system disease; cytosine methylation; antiasthmatic; antiarteriosclerotic; antianaemic; cytostatic; nootropic; neuroprotective; anti-HIV; anticonvulsant; ophthalmological; antirheumatic; antiarthritic; antidiabetic; antipsoriatic; antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia; acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy; neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene; ds.

XX OS Homo sapiens.

XX WO200200928-A2.

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP007537.

XX 30-JUN-2000; 2000DE-01032529.

XX 01-SEP-2000; 2000DE-01043826.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful for diagnosis and treatment of diseases associated with abnormal cytosine methylation.

XX Claim 1; SEQ ID NO 547; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated genes which are modified by the methylation of cytosines. The sequences can be used in the diagnosis and treatment of immune system disorders, including eye diseases such as retinopathy, neovascular glaucoma and macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis, rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel diseases. The present sequence is a gene of the invention

XX SQ Sequence 17959 BP; 4531 A; 360 C; 4258 G; 8710 T; 0 U; 0 Other;

Query Match 80.0%; Score 18.4; DB 6; Length 17959;
Best Local Similarity 95.0%; Pred. NO. 8.3e-02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps

OY 2 ACCTTCACAAAAAATCAA 21
|||||
Db 2483 ACCTTCACAAAAAATCAA 2464
|||||

RESULT 6
ABL54341/c
ID ABL54341 standard; DNA; 17959 BP.
XX AC ABL54341;
XX XX
XX 29-JUL-2002 (first entry)
XX XX
XX Chemically treated apoptosis gene #21.
XX XX
XX Apoptosis; HIV; Bloom syndrome; cardiopathy; neurodegenerative disorder;
KW Herpes simplex virus; renal ischaemia; amyotrophic lateral sclerosis;
XX cancer; ds.
XX XX
XX Unidentified.
XX XX
XX WO200177164-A2.
XX PD 18-OCT-2001.

XX PF 06-APR-2001; 2001WO-EP003969.
XX PF 06-APR-2000; 2000DE-01019058.
XX PR 07-APR-2000; 2000DE-01019173.
XX PR 30-JUN-2000; 2000DE-01032529.
XX PR 01-SEP-2000; 2000DE-01043826.
XX PA (EPIG-) EPIGENOMICS AG.
XX PI Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-017444/02.
XX Chemically modified sequences of genes associated with apoptosis are
PT useful to determine methylation patterns of genomic DNA samples for
PT diagnosis of associated diseases such as cancer.
XX Claim 1; Seq ID #41; 24pp; English.
XX This invention relates to chemically pre-treated DNA of genes associated
CC with apoptosis. The nucleic acids are used to allocate patients for
CC specific therapy for HIV infection, Bloom syndrome, cardiopathy, aging,
CC neurodegenerative disorders, Herpes simplex virus infection, renal
CC ischaemia, ankyrotrophic lateral sclerosis, solid tumours and cancers. This
CC nucleotide sequence represents a chemically treated apoptosis gene. Even
CC SEQ ID numbers are the complementary DNA strands to the odd SEQ ID
CC numbers. The sequence data for this patent is not represented in the
CC printed specification but is based on information supplied by the
CC European patent office
XX SQ Sequence 17959 BP; 4631 A; 360 C; 4258 G; 8710 T; 0 U; 0 Other;
Query Match 80.0%; Score 18.4; DB 6; Length 17959;
Best Local Similarity 95.0%; Pred. No. 9.3e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 ACCTTCACAAAAAAATCAAA 21
DB 2483 ACCITTCACAAAAAAATCAAA 2464
RESULT 7
ABQ76678/c
ID ABQ76678 standard; DNA; 218336 BP.
XX AC ABQ76678;
XX DT 26-MAR-2003 (first entry)
XX DE Androgen receptor signalling pathway-associated DNA AF067844.
XX KW Androgen receptor; transactivation; modulator; Smad3; Smad4; Akt; TGF-B;
KW signal transduction pathway; transforming growth factor-B; phosphatase;
KW tensin; cytosolic; antiproliferative; cellular proliferation; cancer;
KW AF067844; ds.
XX OS Synthetic.
XX WO200282081-A2.
XX PD 17-OCT-2002.
XX PF 05-APR-2002; 2002WO-US011086.
XX PR 06-APR-2001; 2001US-0282266P.
XX PR 13-MAR-2002; 2002US-0365060P.
XX PA (UYRP) UNIV ROCHESTER.
XX PI Chang C;
XX WPI; 2003-046871/04.
XX DR

XX PT Modulating androgen receptor activity, by administering a compound that
PT modulates receptor activity, inhibits receptor-signal transduction
PT pathway/receptor-coactivator interaction or changes amount or receptor.
XX Disclosure; Page 241-299; 302pp; English.
XX This invention describes a novel method for modulating androgen receptor
CC activity or androgen receptor-mediated transactivation activity in a
CC cell. The method involves administering a compound which causes
CC modulation of the androgen receptors activity and the inhibition of
CC interaction between the receptor and a protein involved in a signal
CC transduction pathway. The compound also inhibits the interaction between
CC the androgen receptor and a protein selected from Smad3, Smad4, Akt,
CC transforming growth factor (TGF)-B and phosphatase and tensin homologues
CC deleted on chromosome 10 (PTEN) or their fragments. The compounds of the
CC invention have cytostatic and antiproliferative activity. The obtained
CC composition is useful for treating any disease, where uncontrolled
CC proliferation or cellular proliferation occurs such as cancer, e.g.
CC prostate cancer. This sequence represents the androgen receptor
CC transactivation signalling pathway modulator AF067844 described in the
CC method of the invention
XX SQ Sequence 218336 BP; 64194 A; 39437 C; 43295 G; 71406 T; 0 U; 4 Other;
Query Match 80.0%; Score 18.4; DB 7; Length 218336;
Best Local Similarity 95.0%; Pred. No. 9.3e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 CCTTCACAAAAAAATCAAT 22
DB 58772 CCTTCACAAAAAAATCAAT 58753
RESULT 8
ABK76960
ID ABK76960 standard; DNA; 1506 BP.
XX AC ABK76960;
XX DT 13-AUG-2002 (first entry)
XX DE Bacillus licheniformis genomic sequence tag (GST) #4251.
XX KW Differential gene expression; genomic sequenced tag; GST;
KW altered culture condition; environmental stress;
KW physiological provocation; ds.
XX OS Bacillus licheniformis.
XX PN WO200229113-A2.
XX PD 11-APR-2002.
XX PF 05-OCT-2001; 2001WO-US031437.
XX PR 06-OCT-2000; 2000US-00680598.
XX PR 27-MAR-2001; 2001US-0279526P.
XX PA (NOVO) NOVOZYMES BIOTECH INC.
XX PA (NOVO) NOVOZYMES AS.
XX PI Berka R, Clausen IG;
XX WPI; 2002-416684/44.
XX Monitoring differential expression of several genes in first Bacillus
PT cell relative to expression of same genes in one or more second Bacillus
PT cells, by using substrate containing Bacillus genomic sequenced tag
PT array.
XX Claim 4; SEQ ID NO 4251; 200pp; English.
XX PS

cells in which both alleles of a gene are modified, comprising modifying one allele by insertion or replacement by a cassette having an expressible selectable marker and modifying another allele by recombination, of a promoter replacement fragment with a heterologous promoter, so that expression of the second allele is regulated by the promoter. (M1) is useful for constructing a strain of diploid fungal cells in which both alleles of a gene are modified. The diploid fungal cells having both alleles modified are useful for identifying a gene that is essential to the survival or growth of a fungus, a gene that contributes to the virulence and/or pathogenicity of a fungus, a gene that contributes to the resistance of a diploid fungus to an antifungal agent, an antifungal agent that inhibits the growth of a diploid fungus and for identifying a therapeutic agent for treatment of a mammalian disease. (M2) is useful for identifying a compound which modulates the activity of a gene product, preferably enzymatic activity, carbon compound catabolism, biosynthetic, transporter, transcriptional, translational, signal transduction, DNA replication and cell division activity. The method is useful for identifying a compound having the ability to inhibit growth or proliferation of *C. albicans* cells and for treating infection by *C. albicans*. The present sequence is that of an essential *Candida albicans* gene used in the method of the invention. Note: The sequence data for this patent is not represented in the printed specification but is based on sequence information supplied to Derwent by the European Patent Office.

Sequence 1671 BP; 512 A; 491 C; 282 G; 386 T; 0 U; 0 Other;
 79.1%; Score 18.2; DB 6; Length 1671;
 at Local Similarity 87.0%; Pred. No. 8.7e+03;
 chas 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0

1 CACCTTCACAAAAAATCAATC 23
156 CACCATCAAGATAAATCAATC 178

156 CACCATCACGATATAAATCATC 178

T 10
1142/C
ABL55142 standard; cDNA; 2029 BP.

ABL55142 standard; cDNA; 2029 BP.
ABL55142;
31-MAY-2002 (first entry)

31-MAY-2002 (first entry)
Human NADH dehydrogenase subunit I-10-encoding cDNA.
Human; NADH dehydrogenase subunit I-10; recombinant production;

Human; NADH dehydrogenase subunit I-10; recombinant production;
nicotinamide adenine dinucleotide; cancer; HIV infection;
human immunodeficiency virus; gene therapy; cytostatic; gene;
58.

SS.
Homo sapiens.
Key
CDS
Location/Qualifiers
1452..1730

```

Key      Location/Qualifiers
CDS      1452..1730
          /tag= a
          /product= "Human NADH dehydrogenase subunit I-10"

CN1325978-A.

```

CN1325978-A,
12-DEC-2001,
31-MAY-2000; 2000CN-00116279.

31-MAY-2000; 2000CN-00116279.
31-MAY-2000; 2000CN-00116279.
(BODE-) BODE GENE DEV CO LTD SHANGHAI.

(BODE-) BODE GENE DEV CO LTD SHANGHAI.
Mao Y, Xie Y;
WPI; 2002-196699/26.

WPI; 2002-196699/25.
P-PSDB; AAM49162.

PT Human nicotinamide adenine dinucleotide reduced (NADH) dehydrogenase subunit I-10 and encoding polynucleotide useful for treating cancer and human immunodeficiency virus.

PS Claim 6; Page 26-27 (Disclosure); 34pp; Chinese.

XX This sequence represents cDNA encoding human NADH (nicotinamide adenine dinucleotide) dehydrogenase subunit I-10. The protein has a molecular weight of 10 kD. The invention relates to NADH dehydrogenase subunit I-10 (AAM49162), nucleic acids encoding it (ABL55142), a method for the recombinant production of the protein, an antagonist of the protein, and the use of the protein, gene and antagonist in therapeutic applications.

CC NADH dehydrogenase subunit I-10 can be used in the treatment of a variety of diseases such as cancer and HIV (human immunodeficiency virus) infection

XX Sequence 2029 BP; 556 A; 390 C; 410 G; 672 T; 0 U; 1 Other;

Query Match 79.1%; Score 18.2; DB 6; Length 2029;

Best Local Similarity 87.0%; Pred. No. 8.8e+02;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CACCTTCACAAAAAATCAATC 23

DB 522 CACCTTCACAAAAAATCAATC 500

RESULT 11

ADA20417/c

ID ADA20417 standard; DNA; 5881 BP.

XX

AC ADA20417;

XX

DT 20-NOV-2003 (first entry)

XX

DE Prostate tumour related genomic DNA complement sample #41.

XX

KW cytostatic; gene therapy; genetic marker; epigenetic parameter;

XX

KW classification; differentiation; diagnosis; prostate tumour;

XX

KW prostate cancer; cytosine methylation; uracil;

XX

KW single nucleotide polymorphism; SNP; prostate carcinoma; ss.

XX

OS Homo sapiens.

XX

XX WO2002103042-A2.

PN

XX 27-DEC-2002.

XX

XX 14-JUN-2002; 2002WO-EP006605.

PF

XX 14-JUN-2001; 2001DE-01028508.

XX

XX (EPIG-) EPIGENOMICS AG.

XX

XX Distler J, Model F, Adorjan P;

XX

XX WPI; 2003-167536/16.

DR

XX Determining genetic and/or epigenetic parameters, useful for the classification, differentiation and/or diagnosis of prostate tumors or a predisposition to prostate cancer, comprises analyzing cytosine methylation.

XX

XX Claim 28; Page 310-312; 376pp; English.

XX

CC The invention relates to a method of determining genetic and/or epigenetic parameters for the classification, differentiation and/or diagnosis of prostate tumors or the predisposition to prostate cancer, by analysing cytosine methylation in a sample of genomic DNA. The method comprises chemically treating unmethylated cytosine bases at the 5-position to uracil or another base, which is dissimilar to cytosine in terms of hybridization behaviour; followed by amplifying at least one fragment of the chemically pre-treated genomic DNA using sets of primer

CC oligonucleotides and a polymerase. The oligomers or probes derived from them are useful for detecting the methylation state of all CpG dinucleotides and/or single nucleotide polymorphisms (SNPs) in a chemically pre-treated genomic DNA. They are all useful for treating prostate carcinoma. This sequence represents a fragment of genomic DNA used in the method of the invention.

XX Sequence 5881 BP; 1488 A; 165 C; 1448 G; 2780 T; 0 U; 0 Other;

Query Match 79.1%; Score 18.2; DB 7; Length 5881;

Best Local Similarity 87.0%; Pred. No. 9.3e+02;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CACCTTCACAAAAAATCAATC 23

DB 2215 CACCTTCACAAAAAATCAATC 2193

RESULT 12

ADA84224/c

ID ADA84224 standard; DNA; 5881 BP.

XX

AC ADA84224;

XX

DT 20-NOV-2003 (first entry)

XX

DE Human renal/prostate carcinoma associated DNA SEQ ID NO:82.

XX

KW ds; renal cancer; prostate cancer; cytosine methylation;

XX

KW single nucleotide polymorphism; histological; cytological.

XX

OS Homo sapiens.

XX

XX WO2002103041-A2.

PN

XX 27-DEC-2002.

XX

XX 14-JUN-2002; 2002WO-EP006603.

XX

XX 14-JUN-2001; 2001DE-01028509.

PR

XX (EPIG-) EPIGENOMICS AG.

XX

XX Distler J, Model F, Adorjan P;

XX

XX WPI; 2003-183991/18.

XX

XX Method for characterizing, classifying and/or differentiating renal and prostate cancers, by analyzing the genetic and/or epigenetic parameters of genomic DNA, particularly by determining its cytosine methylation status.

XX

XX Claim 1; Page 176-177; 211pp; English.

XX

CC The invention relates to a novel method for characterising, classifying and/or differentiating renal and prostate cancer. The method comprises extracting genomic DNA from a biological sample, converting cytosine bases (by chemical treatment) that are unmethylated at the 5-position to uracil or another base, and amplifying at least one fragment of the chemically pretreated genomic DNA using sets of primer oligonucleotides and a polymerase. The method is useful for detecting the cytosine methylation state and/or single nucleotide polymorphisms in genomic DNA, particularly for characterising, classifying and/or differentiating renal and prostate cancers. The oligomers are useful as primer oligonucleotides for the amplification of any of the 112 DNA sequences of the invention. The set of oligomer probes is useful for detecting the cytosine methylation state and/or single nucleotide polymorphisms in any of the 112 chemically pretreated genomic DNA sequences. The method is also useful for identifying the tissue of origin of cancer cells. The method allows the classification, differentiation and/or diagnosis of cancer tissues using minute samples which would be inadequate for histological or cytological analysis. The present sequence represents one of the 112 DNA sequences of the invention.

```
XX SQ Sequence 5881 BP; 1488 A; 165 C; 1448 G; 2780 T; 0 U; 0 Other;
Query Match 79.1%; Score 18.2; DB 7; Length 5881;
Best Local Similarity 87.0%; Pred. No. 9.3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CACCTTCACAAAAAATCAATC 23
Db 2215 CACCTCGATAAAAAAATCAATC 2193

RESULT 13
ABL33031/c
ID ABL33031 standard; DNA; 6118 BP.
XX ABL33031;
AC ABL33031;
XX 26-MAR-2002 (first entry)
XX Human immune system associated gene SEQ ID NO: 1004.
XX Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytosatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
XX ds.
XX Homo sapiens.
XX OS
XX WO200200928-A2.
XX PN
XX 03-JAN-2002.
XX PD
XX 02-JUL-2001; 2001WO-EP007537.
XX PF
XX 30-JUN-2000; 2000DE-01032529.
XX PR
XX 01-SEP-2000; 2000DE-01043826.
XX PA (EPiG-) EPIGENOMICS AG.
XX PI Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-130909/17.
XX DR
XX Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX PT
XX Claim 1; SEQ ID NO 1004; 32pp + Sequence Listing; German.
XX PS
XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX CC
SQ Sequence 6118 BP; 1751 A; 141 C; 1229 G; 2997 T; 0 U; 0 Other;
Query Match 79.1%; Score 18.2; DB 6; Length 6118;
Best Local Similarity 87.0%; Pred. No. 9.3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CACCTTCACAAAAAATCAATC 23
Db 1016 CACCTTCACAAAAAATCAATC 994
```

```
RESULT 14
ABL33687/c
ID ABL33687 standard; DNA; 8781 BP.
XX ABL33687;
AC ABL33687;
XX 26-MAR-2002 (first entry)
XX Human immune system associated gene SEQ ID NO: 1660.
XX Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytosatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
XX ds.
XX Homo sapiens.
XX OS
XX WO200200928-A2.
XX PN
XX 03-JAN-2002.
XX PD
XX 02-JUL-2001; 2001WO-EP007537.
XX PF
XX 30-JUN-2000; 2000DE-01032529.
XX PR
XX 01-SEP-2000; 2000DE-01043826.
XX PA (EPiG-) EPIGENOMICS AG.
XX PI Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-130909/17.
XX DR
XX Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX PT
XX Claim 1; SEQ ID NO 1660; 32pp + Sequence Listing; German.
XX PS
XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX CC
SQ Sequence 8781 BP; 2159 A; 65 C; 2122 G; 4435 T; 0 U; 0 Other;
Query Match 79.1%; Score 18.2; DB 6; Length 8781;
Best Local Similarity 87.0%; Pred. No. 9.5e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CACCTTCACAAAAAATCAATC 23
Db 8498 CACCTTCACAAAAAATCAATC 8476

RESULT 15
ABL33859/c
ID ABL33859 standard; DNA; 13326 BP.
XX ABL33859;
AC ABL33859;
XX 26-MAR-2002 (first entry)
XX Human immune system associated gene SEQ ID NO: 1832.
XX
```

KW	Human; immune system disease; cytosine methylation; antiasthmatic;
KW	antiarteriosclerotic; anianaemic; cytosatic; neotropic;
KW	neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW	antirheumatic; antiarthritis; antidiabetic; antipsoriatic;
KW	antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW	acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW	neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
ds.	
XX	
XX	Homo sapiens.
OS	
XX	WO200200928-A2.
PN	
XX	
PD	03-JAN-2002.
XX	
PF	02-JUL-2001; 2001WO-EP007537.
PP	
PR	30-JUN-2000; 2000DE-01032529.
XX	
PR	01-SEP-2000; 2000DE-01043826.
XX	
XX	(EPIG-) EPIGENOMICS AG.
PA	
XX	
PI	Olek A, Piepenbrock C, Berlin K;
XX	
DR	WPI; 2002-130909/17.
XX	
PT	Nucleic acid comprising fragment of chemically modified gene, useful for
PT	diagnosis and treatment of diseases associated with abnormal cytosine
PT	methylation.
XX	
XX	Claim 1; SEQ ID NO 1832; 32pp + Sequence Listing; German.
PS	
XX	
CC	The present invention provides a number of human immune system associate
CC	genes which are modified by the methylation of cytosines. The sequences
CC	can be used in the diagnosis and treatment of immune system disorders,
CC	including eye diseases such as retinopathy, neovascular glaucoma and,
CC	macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC	leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC	rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC	diseases. The present sequence is a gene of the invention
XX	
XX	Sequence 13326 BP; 3317 A; 258 C; 3320 G; 6431 T; 0 U; 0 Other;
SQ	
Query Match	79.1%; Score 18.2; DB 6; Length 13326;
Best Local Similarity	87.0%; Pred.No. 9.7e+02;
Matches	20; Conservative 0; Mismatches 3; Indels 0; Gaps
Qy	1 CACCTTCACAAAAAATCAATC 23
Db	6504 CATCTTCCCAAAATAATCAATC 6482

Search completed: April 24, 2004, 18:45:02
Job time : 197.506 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2004, 18:26:15 ; Search time 42.5926 Seconds

(without alignments)
299.673 Million cell updates/sec

Title: US-10-084-555A-116

Perfect score: 23

Sequence: 1 cacccttcacaaaataatcatc 23

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18.2	79.1	549	4	US-09-543-681A-2149
2	17.8	77.4	580073	4	US-08-545-528D-1
3	17.4	75.7	1097	4	US-09-976-594-115
4	17.2	74.8	282	4	US-09-328-352-726
5	17.2	74.8	354	4	US-09-134-000C-2933
6	17.2	74.8	486	4	US-09-134-001C-1428
7	17.2	74.8	711	4	US-09-134-000C-3250
8	17.2	74.8	786	4	US-09-780-988-2
9	17.2	74.8	9388	3	US-08-991-789A-141
10	17.2	74.8	9388	4	US-09-062-451-141
11	17.2	74.8	9388	4	US-09-598-326-141
12	17.2	74.8	9388	4	US-09-289-198-141
13	17.2	74.8	9388	4	US-09-429-755-141
14	17.2	74.8	319608	4	US-09-539-333D-1
15	17.2	74.8	319608	4	US-09-692-409-1
16	17.2	74.8	1684976	4	US-08-916-421B-1
17	16.8	73.0	260	3	US-08-134-198E-11
18	16.8	73.0	278	4	US-09-313-294A-2339
19	16.8	73.0	789	4	US-09-134-000C-2131
20	16.8	73.0	1490	1	US-08-471-601-1
21	16.8	73.0	1490	1	US-08-474-556-1
22	16.8	73.0	1490	1	US-08-351-899-1
23	16.8	73.0	1490	1	US-08-479-382-1
24	16.8	73.0	1490	1	US-08-479-354-1
25	16.8	73.0	1490	1	US-08-479-383-1
26	16.8	73.0	1490	2	US-08-479-041-1
27	16.8	73.0	1490	3	US-08-819-646-1

28 16.8 73.0 1490 4 US-09-195-716-1 Sequence 1, Appli
 29 16.8 73.0 10302 4 US-09-657-960-1 Sequence 1, Appli
 30 16.8 73.0 23356 4 US-09-976-594-115 Sequence 115, App
 31 16.6 72.2 294 4 US-09-313-294A-6560 Sequence 6560, Ap
 32 16.6 72.2 3182 1 US-07-797-556-5 Sequence 5, Appli
 33 16.6 72.2 3182 1 US-07-943-843-1 Sequence 1, Appli
 34 16.6 72.2 3182 1 US-08-347-003-1 Sequence 1, Appli
 35 16.6 72.2 3591 1 US-07-943-843-5 Sequence 5, Appli
 36 16.6 72.2 3591 1 US-08-347-003-5 Sequence 5, Appli
 37 16.6 72.2 5496 3 US-08-600-982-23 Sequence 23, Appl
 38 16.6 72.2 5496 5 PCT-US94-10261A-23 Sequence 23, Appl
 39 16.6 72.2 5844 4 US-10-204-708-90 Sequence 90, Appl
 40 16.6 72.2 31096 4 US-08-956-171E-59 Sequence 59, Appl
 41 16.4 71.3 3543 3 US-09-224-024-27 Sequence 27, Appl
 42 16.4 71.3 3543 5 PCT-US94-07902-27 Sequence 27, Appl
 43 16.4 71.3 392000 4 US-10-027-983-11 Sequence 11, Appl
 44 16.2 70.4 25 3 US-08-781-891-108 Sequence 108, App
 45 16.2 70.4 25 4 US-09-618-166-108 Sequence 108, App

ALIGNMENTS

RESULT 1

US-09-543-681A-2149
 ; Sequence 2149, Application US/09543681A
 ; Patent No. 6605709
 ; GENERAL INFORMATION:
 ; APPLICANT: GARY BRETON
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABI
 ; FILE REFERENCE: 2709 1002-001
 ; CURRENT APPLICATION NUMBER: US/09/543,681A
 ; PRIOR FILING DATE: 2000-04-05
 ; PRIOR APPLICATION NUMBER: US 60/128,706
 ; PRIOR FILING DATE: 1999-04-09
 ; NUMBER OF SEQ ID NOS: 8344
 ; SEQ ID NO 2149
 ; LENGTH: 549
 ; TYPE: DNA
 ; ORGANISM: Proteus mirabilis
 ; US-09-543-681A-2149

Query Match 79.1%; Score 18.2; DB 4; Length 549;
 Best Local Similarity 87.0%; Pred. No. 99;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CACCTTCACAAAAAATCATC 23
 DB 40 CACCTTCACATATAATAATC 62

RESULT 2

US-08-545-528D-1/C
 ; Sequence 1, Application US/08545528D
 ; Patent No. 6537773
 ; GENERAL INFORMATION:
 ; APPLICANT: Fraser et al.
 ; TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragment
 ; Patent No. 6537773
 ; TITLE OF INVENTION: Thereof, and Uses Thereof
 ; FILE REFERENCE: PB193PI
 ; CURRENT APPLICATION NUMBER: US/08/545,528D
 ; PRIOR FILING DATE: 1995-10-19
 ; PRIOR APPLICATION NUMBER: US 08/488,018
 ; PRIOR FILING DATE: 1995-06-07
 ; PRIOR APPLICATION NUMBER: US 08/473,545
 ; NUMBER OF SEQ ID NOS: 1
 ; SOFTWARE: Patent version 3.1
 ; SEQ ID NO 1
 ; LENGTH: 580073
 ; TYPE: DNA

```

; ORGANISM: Mycoplasma genitalium
US-08-545-528D-1

Query Match      77.4%; Score 17.8; DB 4; Length 580073;
Best Local Similarity 90.5%; Pred. No. 2e+02; 2; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 0;

QY 1 CACCTTCACAAAAAATCAA 21
   |||||
DB 425081 CACTTTCACAAATAAATCAA 425061

RESULT 3
US-09-976-594-115
; Sequence 115, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 115
; LENGTH: 1097
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 4349106CB1
US-09-976-594-115

Query Match      75.7%; Score 17.4; DB 4; Length 1097;
Best Local Similarity 94.7%; Pred. No. 2.1e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CACCTTCACAAAAAATC 19
   |||||
DB 741 CATCTTCACAAAAAATC 759

RESULT 4
US-09-328-352-726
; Sequence 726, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 726
; LENGTH: 282
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-726

Query Match      74.8%; Score 17.2; DB 4; Length 282;
Best Local Similarity 86.4%; Pred. No. 2.3e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ACCTTCACAAAAAATCAATC 23
   |||||
DB 114 ACTTTCACAAAAAATCAATC 135

RESULT 5

```

```

US-09-134-000C-2933/c
; Sequence 2933, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2933
; LENGTH: 354
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-134-000C-2933

Query Match      74.8%; Score 17.2; DB 4; Length 354;
Best Local Similarity 86.4%; Pred. No. 2.3e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CACCTTCACAAAAAATCAAT 22
   |||||
DB 103 CACCTTCACATCAAAATCAAT 82

RESULT 6
US-09-134-001C-1428/c
; Sequence 1428, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1428
; LENGTH: 486
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1428

Query Match      74.8%; Score 17.2; DB 4; Length 486;
Best Local Similarity 86.4%; Pred. No. 2.4e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CACCTTCACAAAAAATCAAT 22
   |||||
DB 349 CACCTTCACGTAGAAATCAAT 328

RESULT 7
US-09-134-000C-3250
; Sequence 3250, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15

```

; NUMBER OF SEQ ID NOS: 6812
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 3250
 ; LENGTH: 711
 ; TYPE: DNA
 ; ORGANISM: Enterococcus faecalis
 US-09-134-000C-3250
 Query Match 74.8%; Score 17.2; DB 4; Length 711;
 Best Local Similarity 86.4%; Pred. No. 2.5e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CACCTTCACAAAAAATCAAT 22
 Db 257 CAACTTCACCTTAAAAAATCAAT 278

RESULT 8

US-09-790-988-2/c
 ; Sequence 2, Application US/09790988
 ; Patent No. 6632935

; GENERAL INFORMATION:
 ; APPLICANT: SHIGENOBU, SHUJI
 ; APPLICANT: WATANABE, HIDEKI
 ; APPLICANT: HATTORI, MASAHIRO
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
 ; FILE REFERENCE: 081356/0159
 ; CURRENT APPLICATION NUMBER: US/09/790,988
 ; CURRENT FILING DATE: 2001-02-23
 ; PRIOR APPLICATION NUMBER: JP2000-107160
 ; PRIOR FILING DATE: 2000-04-07
 ; NUMBER OF SEQ ID NOS: 7
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 2
 ; LENGTH: 7786
 ; TYPE: DNA
 ; ORGANISM: Buchnera sp.
 US-09-790-988-2

Query Match 74.8%; Score 17.2; DB 4; Length 7786;
 Best Local Similarity 86.4%; Pred. No. 2.9e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CACCTTCACAAAAAATCAAT 22
 Db 3458 CATCTTCACAGAAATTCAT 3437

RESULT 9

US-08-991-789A-141/c
 ; Sequence 141, Application US/08991789A
 ; Patent No. 6225054

; GENERAL INFORMATION:
 ; APPLICANT: Frudakis, Tony N.
 ; APPLICANT: Smith, John M.
 ; APPLICANT: Reed, Steven G.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
 ; TREATMENT AND DIAGNOSIS OF BREAST CANCER
 ; NUMBER OF SEQUENCES: 292
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Seed IP Law Group
 ; STREET: 701 Fifth Avenue, Suite 6300
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: IBM PC compatible
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/991,789A
 ; FILING DATE: 11-Dec-1997
 ; CLASSIFICATION: <Unknown>
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Potter, Jane E. R.
 ; REGISTRATION NUMBER: 33,332
 ; REFERENCE/DOCKET NUMBER: 210121.419C3
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 141:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 9388 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 141:
 US-08-991-789A-141

Query Match 74.8%; Score 17.2; DB 3; Length 9388;
 Best Local Similarity 86.4%; Pred. No. 2.9e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CACCTTCACAAAAAATCAAT 22
 Db 8466 CTCCTTCACAAAAAATAAT 8445

RESULT 10

US-09-062-451-141/c
 ; Sequence 141, Application US/09062451
 ; Patent No. 6344550

; GENERAL INFORMATION:
 ; APPLICANT: Frudakis, Tony N.
 ; APPLICANT: Smith, John M.
 ; APPLICANT: Reed, Steven G.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
 ; TREATMENT AND DIAGNOSIS OF BREAST CANCER
 ; NUMBER OF SEQUENCES: 297
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED and BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA: US/09/062,451
 ; FILING DATE: 04-APR-1997
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Maki, David J.
 ; REGISTRATION NUMBER: 31,392
 ; REFERENCE/DOCKET NUMBER: 210121.419C2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 141:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 9388 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 141:
 US-09-062-451-141

Query Match 74.8%; Score 17.2; DB 4; Length 9388;
 Best Local Similarity 86.4%; Pred. No. 2.9e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CACCTTCACAAAAAATCAAT 22
Db 8466 CTCCTTCTCAAAAAAATAAT 8445

RESULT 11

US-09-598-326-141/c
; Sequence 141, Application US/09598326
; Patent No. 6423496
; GENERAL INFORMATION:
; APPLICANT: Fridakis, Tony N.
; Smith, John M.
; Reed, Steven G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF BREAST CANCER
; NUMBER OF SEQUENCES: 247
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed Intellectual Property Law Group PLLC
; STREET: 701 Fifth Avenue, Suite 6300
; City: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/598,326
; FILING DATE: 20-Jun-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Potter, Jane E.R.
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 210121.419D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 141:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9388 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 141:
US-09-598-326-141

Query Match 74.8%; Score 17.2; DB 4; Length 9388;
Best Local Similarity 86.4%; Pred. No. 2.9e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CACCTTCACAAAAAATCAAT 22
Db 8466 CTCCTTCTCAAAAAAATAAT 8445

RESULT 12

US-09-289-198-141/c
; Sequence 141, Application US/09289198
; Patent No. 6586570
; GENERAL INFORMATION:
; APPLICANT: Fridakis, Tony N.
; Smith, John M.
; Reed, Steven G.
; APPLICANT: Mishner, Lynda
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.419C5
; CURRENT APPLICATION NUMBER: US/09/289,198
; CURRENT FILING DATE: 1999-04-09
; EARLIER APPLICATION NUMBER: US 09/062,451

; EARLIER FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: US 08/991,789
; EARLIER FILING DATE: 1997-12-11
; EARLIER APPLICATION NUMBER: US 08/838,762
; EARLIER FILING DATE: 1997-04-09
; EARLIER APPLICATION NUMBER: PCT/US97/00485
; EARLIER FILING DATE: 1997-01-10
; EARLIER APPLICATION NUMBER: US 08/700,014
; EARLIER FILING DATE: 1996-08-20
; EARLIER APPLICATION NUMBER: US 08/585,392
; EARLIER FILING DATE: 1996-01-01
; NUMBER OF SEQ ID NOS: 312
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 141
; LENGTH: 9388
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-289-198-141

Query Match 74.8%; Score 17.2; DB 4; Length 9388;
Best Local Similarity 86.4%; Pred. No. 2.9e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CACCTTCACAAAAAATCAAT 22
Db 8466 CTCCTTCTCAAAAAAATAAT 8445

RESULT 13

US-09-429-755-141/c
; Sequence 141, Application US/09429755A
; Patent No. 6656480
; GENERAL INFORMATION:
; APPLICANT: Fridakis, Tony N.
; Smith, John M.
; APPLICANT: Reed, Steven G.
; APPLICANT: Mishner, Lynda
; APPLICANT: Retter, Marc W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.419C6
; CURRENT APPLICATION NUMBER: US/09/429,755A
; CURRENT FILING DATE: 1999-10-28
; NUMBER OF SEQ ID NOS: 315
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 141
; LENGTH: 9388
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-429-755-141

Query Match 74.8%; Score 17.2; DB 4; Length 9388;
Best Local Similarity 86.4%; Pred. No. 2.9e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CACCTTCACAAAAAATCAAT 22
Db 8466 CTCCTTCTCAAAAAAATAAT 8445

RESULT 14

US-09-539-333D-1/c
; Sequence 1, Application US/09539333D
; Patent No. 6476208
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Chumakov, Ilya
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Bihain, Bernard
; APPLICANT: Essieux, Laurent
; TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS

FILE REFERENCE: GENSET.04TAUS
CURRENT APPLICATION NUMBER: US/09/539,333D
CURRENT FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: US 60/126,903
PRIOR FILING DATE: 1999-03-30
PRIOR APPLICATION NUMBER: US 60/131,971
PRIOR FILING DATE: 1999-04-30
PRIOR APPLICATION NUMBER: US 60/132,065
PRIOR FILING DATE: 1999-04-30
PRIOR APPLICATION NUMBER: US 60/143,928
PRIOR FILING DATE: 1999-07-14
PRIOR APPLICATION NUMBER: US 60/145,915
PRIOR FILING DATE: 1999-07-27
PRIOR APPLICATION NUMBER: US 60/146,453
PRIOR FILING DATE: 1999-07-29
PRIOR APPLICATION NUMBER: US 60/146,452
PRIOR FILING DATE: 1999-07-29
PRIOR APPLICATION NUMBER: US 60/162,288
PRIOR FILING DATE: 1999-10-28
PRIOR APPLICATION NUMBER: US 09/416,384
PRIOR FILING DATE: 1999-10-12
NUMBER OF SEQ ID NOS: 231
SOFTWARE: Patent.pm
SEQ ID NO 1
LENGTH: 319608
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: 31..1107
OTHER INFORMATION: 5'regulatory region g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 1108..1289
OTHER INFORMATION: exon A g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 14877..14920
OTHER INFORMATION: exon B g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 18778..18862
OTHER INFORMATION: exon Bbis g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 25593..25740
OTHER INFORMATION: exon C g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 29388..29502
OTHER INFORMATION: exon D g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 29967..30282
OTHER INFORMATION: exon E g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 64656..64812
OTHER INFORMATION: exon F g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 65505..65853
OTHER INFORMATION: exon G g35018 gene
FEATURE:
NAME/KEY: misc feature
LOCATION: 65854..67854
OTHER INFORMATION: 3'regulatory region g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 94124..94964
OTHER INFORMATION: exon g35017
FEATURE:
NAME/KEY: exon

LOCATION: 201188..201234
OTHER INFORMATION: exon S g35030 gene
FEATURE:
NAME/KEY: exon
LOCATION: 214676..214793
OTHER INFORMATION: exon T g35030 gene
FEATURE:
NAME/KEY: exon
LOCATION: 215702..215746
OTHER INFORMATION: exon U g35030 gene
FEATURE:
NAME/KEY: exon
LOCATION: 216836..216915
OTHER INFORMATION: exon V g35030 gene
FEATURE:
NAME/KEY: misc feature
LOCATION: 213818..215818
OTHER INFORMATION: 3'regulatory region g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 215819..215941
OTHER INFORMATION: exon R complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 215819..215975
OTHER INFORMATION: exon Rbis complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 216661..216952
OTHER INFORMATION: exon Qbis complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 216661..217061
OTHER INFORMATION: exon Q complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 217027..217061
OTHER INFORMATION: exon Q1 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 229647..229742
OTHER INFORMATION: exon X complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 230408..230721
OTHER INFORMATION: exon P complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 231272..231412
OTHER INFORMATION: exon Obis complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 231787..231880
OTHER INFORMATION: exon O2 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 231870..231879
OTHER INFORMATION: exon O1 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 234174..234321
OTHER INFORMATION: exon Q complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 237406..237428
OTHER INFORMATION: exon Nbis complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 239719..239807
OTHER INFORMATION: exon N2 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 239719..239853


```
OTHER INFORMATION: exon N complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240528..240569
OTHER INFORMATION: exon M117 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240528..240596
OTHER INFORMATION: exon M1090 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240528..240617
OTHER INFORMATION: exon M1069 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240528..240644
OTHER INFORMATION: exon M52 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240528..240824
OTHER INFORMATION: exon M862 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240528..240994
OTHER INFORMATION: exon M692 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240528..241685
OTHER INFORMATION: exon M1 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240800..240993
OTHER INFORMATION: exon M51 complement g34872 gene
FEATURE:
NAME/KEY: misc feature
LOCATION: 241686..243685
OTHER INFORMATION: 5'regulatory region g34872 gene
FEATURE:
NAME/KEY: misc feature
LOCATION: 290652..292852
OTHER INFORMATION: 3'regulatory region g34665 gene
FEATURE:
NAME/KEY: exon
LOCATION: 292653..292841
Query Match 74.8%; Score 17.2; DB 4; Length 319608;
Best Local Similarity 86.4%; Pred. No. 3 3e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CACCTTCACAAAATAATCAAT 22
Db 103043 CACCTTCACAGTAAATCAAT 103022

RESULT 15
US-09-679-409-1/c
Sequence 1, Application US/09679409
Patent No. 6555316
GENERAL INFORMATION:
APPLICANT: Cohen, Daniel
APPLICANT: Blumenfeld, Marta
APPLICANT: Chumakov, Ilya
APPLICANT: Bougueleret, Lydie
APPLICANT: Bessieux, Laurent
TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENE, PROTEINS AND BIALLELIC MARKERS
FILE REFERENCE: 53.US15.CIP
CURRENT APPLICATION NUMBER: US/09/679,409
CURRENT FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 09/539,333
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: 09/416,384
PRIOR FILING DATE: 1999-10-12
PRIOR APPLICATION NUMBER: 60/168,088
```

```
PRIOR FILING DATE: 1999-11-30
NUMBER OF SEQ ID NOS: 134
SOFTWARE: Patent.pm
SEQ ID NO. 1
LENGTH: 319608
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: 199122..201122
OTHER INFORMATION: 5'regulatory region
NAME/KEY: exon
LOCATION: 201123..201234
OTHER INFORMATION: exon S
NAME/KEY: exon
LOCATION: 201123..201560
OTHER INFORMATION: exon S2
NAME/KEY: exon
LOCATION: 214676..214793
OTHER INFORMATION: exon T
NAME/KEY: exon
LOCATION: 215702..215746
OTHER INFORMATION: exon U
NAME/KEY: exon
LOCATION: 216836..216994
OTHER INFORMATION: exon V
NAME/KEY: exon
LOCATION: 216836..217077
OTHER INFORMATION: exon V1
NAME/KEY: exon
LOCATION: 217671..217764
OTHER INFORMATION: exon V1
NAME/KEY: exon
LOCATION: 227655..227736
OTHER INFORMATION: exon V4
NAME/KEY: exon
LOCATION: 238715..238919
OTHER INFORMATION: exon V3
NAME/KEY: exon
LOCATION: 240440..240673
OTHER INFORMATION: exon W
NAME/KEY: exon
LOCATION: 240440..241153
OTHER INFORMATION: exon W2
NAME/KEY: exon
LOCATION: 241072..241291
OTHER INFORMATION: exon X
NAME/KEY: exon
LOCATION: 244353..244561
OTHER INFORMATION: exon Y
NAME/KEY: exon
LOCATION: 246273..247802
OTHER INFORMATION: exon Z
NAME/KEY: misc feature
LOCATION: 247803..249803
OTHER INFORMATION: 3'regulatory region
NAME/KEY: allele
LOCATION: 8316
OTHER INFORMATION: 99-27943-150 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 21672
OTHER INFORMATION: 99-27935-193 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 65485
OTHER INFORMATION: 8-128-33 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 95396
OTHER INFORMATION: 99-31960-363 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 107281
OTHER INFORMATION: 99-24656-260 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 160640
```

OTHER INFORMATION: 99-24639-163 : polymorphic base A or C
NAME/KEY: allele
LOCATION: 160876
OTHER INFORMATION: 99-24634-108 : polymorphic base A or T
NAME/KEY: allele
LOCATION: 168974
OTHER INFORMATION: 99-7652-162 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 170810
OTHER INFORMATION: 99-16100-147 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 173358
OTHER INFORMATION: 99-5862-167 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 189957
OTHER INFORMATION: 99-5919-215 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 197163
OTHER INFORMATION: 99-24658-410 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 200778
OTHER INFORMATION: 8-303-235 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 202651
OTHER INFORMATION: 8-300-221 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 202679
OTHER INFORMATION: 8-300-193 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 203378
OTHER INFORMATION: 8-299-128 : polymorphic base A or T
NAME/KEY: allele
LOCATION: 204138
OTHER INFORMATION: 8-296-213 : polymorphic base A or T
NAME/KEY: allele
LOCATION: 204605
OTHER INFORMATION: 8-252-190 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 204934
OTHER INFORMATION: 99-24644-194 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 205206
OTHER INFORMATION: 8-295-248 : polymorphic base A or C
NAME/KEY: allele
LOCATION: 205329
OTHER INFORMATION: 8-295-125 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 206084
OTHER INFORMATION: 8-293-130 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 206545
OTHER INFORMATION: 8-292-198 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 207313
OTHER INFORMATION: 8-251-322 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 208285
OTHER INFORMATION: 8-289-322 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 208960
OTHER INFORMATION: 8-287-249 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 209123
OTHER INFORMATION: 8-287-86 : polymorphic base A or T
NAME/KEY: allele
LOCATION: 209631
OTHER INFORMATION: 8-285-319 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 210361
OTHER INFORMATION: 8-283-278 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 210463
OTHER INFORMATION: 8-283-176 : polymorphic base A or G

NAME/KEY: allele
LOCATION: 210486
OTHER INFORMATION: 8-283-153 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 210583
OTHER INFORMATION: 8-283-56 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 210879
OTHER INFORMATION: 8-282-345 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 210964
OTHER INFORMATION: 8-282-260 : polymorphic base G or T
NAME/KEY: allele
LOCATION: 210979
OTHER INFORMATION: 8-282-245 : polymorphic base A or C
NAME/KEY: allele
LOCATION: 211050
OTHER INFORMATION: 8-282-174 : variable motif AAAGG or GAAGGAAGGAAGGAAGA
NAME/KEY: allele
LOCATION: 211132
OTHER INFORMATION: 8-282-92 : polymorphic base A or T
NAME/KEY: allele
LOCATION: 211247
OTHER INFORMATION: 8-281-367 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 211315
OTHER INFORMATION: 8-281-299 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 211366
OTHER INFORMATION: 8-281-248 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 212520
OTHER INFORMATION: 8-279-197 : polymorphic base A or C
NAME/KEY: allele
LOCATION: 212821
OTHER INFORMATION: 8-278-289 : polymorphic base C or T
NAME/KEY: allele

Query Match 74.8%; Score 17.2; DB 4; Length 319608;
Best Local Similarity 86.4%; Pred. No. 3.3e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CACCTTCACAAAAAATCAAT 22
Db 103043 CACCTTCACAGTAAAAATCATT 103022

Search completed: April 24, 2004, 21:13:36
Job time : 45.5926 secs

Query Match 81.7%; Score 18.8; DB 13; Length 820;
Best Local Similarity 90.9%; Pred. No. 7.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ACCTTCACAAAAAAATCAA 21
 |||||
 Db 5375 ACCTTCATAAAAAAATCAA 5356

RESULT 6

US-10-311-455-1171/c
 ; Sequence 1171, Application US/10311455
 ; Publication No. US20030143606A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OLEK, Alexander
 ; APPLICANT: PIEPENBROCK, Christian
 ; APPLICANT: BERLIN, Kurt
 ; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determination of Cytosine Methylation
 ; FILE REFERENCE: 5013.1014
 ; CURRENT APPLICATION NUMBER: US/10/311,455
 ; PRIOR FILING DATE: 2002-12-16
 ; PRIOR APPLICATION NUMBER: PCT/EP01/07537
 ; PRIOR FILING DATE: 2001-07-02
 ; PRIOR APPLICATION NUMBER: DE 10032529.7
 ; PRIOR FILING DATE: 2000-06-30
 ; PRIOR APPLICATION NUMBER: DE 10043826.1
 ; PRIOR FILING DATE: 2000-09-01
 ; NUMBER OF SEQ ID NOS: 2424
 ; SEQ ID NO 1171
 ; LENGTH: 7327
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
 US-10-311-455-1171

Query Match 80.0%; Score 18.4; DB 15; Length 7327;
 Best Local Similarity 95.0%; Pred. No. 1.6e+03;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ACCTTCACAAAAAAATCAA 21
 |||||
 Db 4178 ACCTTCATAAAAAAATCAA 4159

RESULT 7

US-10-311-455-547/c
 ; Sequence 547, Application US/10311455
 ; Publication No. US20030143606A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OLEK, Alexander
 ; APPLICANT: PIEPENBROCK, Christian
 ; APPLICANT: BERLIN, Kurt
 ; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determination of Cytosine Methylation
 ; FILE REFERENCE: 5013.1014
 ; CURRENT APPLICATION NUMBER: US/10/311,455
 ; PRIOR FILING DATE: 2002-12-16
 ; PRIOR APPLICATION NUMBER: PCT/EP01/07537
 ; PRIOR FILING DATE: 2001-07-02
 ; PRIOR APPLICATION NUMBER: DE 10032529.7
 ; PRIOR FILING DATE: 2000-06-30
 ; PRIOR APPLICATION NUMBER: DE 10043826.1
 ; PRIOR FILING DATE: 2000-09-01
 ; NUMBER OF SEQ ID NOS: 2424
 ; SEQ ID NO 547
 ; LENGTH: 17959
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
 US-10-311-455-547

Query Match 80.0%; Score 18.4; DB 15; Length 17959;

Best Local Similarity 95.0%; Pred. No. 1.8e+03;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ACCTTCACAAAAAAATCAA 21
 |||||
 Db 2483 ACCTTCATAAAAAAATCAA 2464

RESULT 8

US-10-240-452-41/c
 ; Sequence 41, Application US/10240452
 ; Publication No. US20030162194A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OLEK, Alexander
 ; APPLICANT: PIEPENBROCK, Christian
 ; APPLICANT: BERLIN, Kurt
 ; TITLE OF INVENTION: Diagnosis of Diseases Associated with Apoptosis
 ; FILE REFERENCE: 5013.1006
 ; CURRENT APPLICATION NUMBER: US/10/240,452
 ; PRIOR FILING DATE: 2002-10-02
 ; PRIOR APPLICATION NUMBER: PCT/EP01/03969
 ; PRIOR FILING DATE: 2001-04-06
 ; PRIOR APPLICATION NUMBER: DE 10019058.8
 ; PRIOR FILING DATE: 2000-04-06
 ; PRIOR APPLICATION NUMBER: DE 10019173.8
 ; PRIOR FILING DATE: 2000-04-07
 ; PRIOR APPLICATION NUMBER: DE 10032529.7
 ; PRIOR FILING DATE: 2000-06-30
 ; PRIOR APPLICATION NUMBER: DE 10043826.1
 ; PRIOR FILING DATE: 2000-09-01
 ; NUMBER OF SEQ ID NOS: 78
 ; SEQ ID NO 41
 ; LENGTH: 17959
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
 US-10-240-452-41

Query Match 80.0%; Score 18.4; DB 15; Length 17959;
 Best Local Similarity 95.0%; Pred. No. 1.8e+03;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ACCTTCACAAAAAAATCAA 21
 |||||
 Db 2483 ACCTTCATAAAAAAATCAA 2464

RESULT 9

US-10-312-841-1/c
 ; Sequence 1, Application US/10312841
 ; Publication No. US20030186277A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Epigenomics AG
 ; TITLE OF INVENTION: Diagnose von bedeutenenden genetischen Parametern innerhalb des MH
 ; FILE REFERENCE: E01/1208/WO
 ; CURRENT APPLICATION NUMBER: US/10/312,841
 ; CURRENT FILING DATE: 2002-12-30
 ; NUMBER OF SEQ ID NOS: 2
 ; SEQ ID NO 1
 ; LENGTH: 3673778
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
 ; FEATURE:
 ; NAME/KEY: unsure
 ; LOCATION: (3294164)
 US-10-312-841-1

Query Match 80.0%; Score 18.4; DB 15; Length 3673778;
 Best Local Similarity 95.0%; Pred. No. 1.2e+03;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

US-10-027-632-230654

Query Match      79.1%; Score 18.2; DB 13; Length 548;
Best Local Similarity 87.0%; Pred. No. 1.2e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1  CACCTTCACAAAAAAATCAATC 23
          |||||
Db      382 CCCTTTCTCAAAAAAATCAATC 360

RESULT 12
US-10-027-632-230654/c
; Sequence 230654, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIORITY APPLICATION NUMBER: US 60/218,006
; PRIORITY FILING DATE: 2000-07-12
; PRIORITY APPLICATION NUMBER: US 60/198,676
; PRIORITY FILING DATE: 2000-04-20
; PRIORITY APPLICATION NUMBER: US 60/193,483
; PRIORITY FILING DATE: 2000-03-29
; PRIORITY APPLICATION NUMBER: US 60/185,218
; PRIORITY FILING DATE: 2000-02-24
; PRIORITY APPLICATION NUMBER: US 60/167,363
; PRIORITY FILING DATE: 1999-11-23
; PRIORITY APPLICATION NUMBER: US 60/156,358
; PRIORITY FILING DATE: 1999-09-28
; PRIORITY APPLICATION NUMBER: US 60/146,002
; PRIORITY FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 230654
; LENGTH: 548
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(548)
; OTHER INFORMATION: n = A, T, C or G

US-10-027-632-230654

Query Match      79.1%; Score 18.2; DB 16; Length 548;
Best Local Similarity 87.0%; Pred. No. 1.2e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1  CACCTTCACAAAAAAATCAATC 23
          |||||
Db      382 CCCTTTCTCAAAAAAATCAATC 360

RESULT 13
US-10-425-114-34439/c
; Sequence 34439, Application US/10425114
; Publication No. US2004003488A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28

```

; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 34439
; LENGTH: 1024
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLMO17267B12_FLI
US-10-425-114-34439

Query Match 79.1%; Score 18.2; DB 13; Length 1024;
Best Local Similarity 87.0%; Pred. No. 1.3e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CACCTTCACAAACAAATCAATC 23
|||||
Db 624 CACCTTCACAAATCAATC 602

RESULT 14

US-10-425-114-28487/c
; Sequence 28487, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 28487
; LENGTH: 1081
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB4758-006-B8_FLI
US-10-425-114-28487

Query Match 79.1%; Score 18.2; DB 13; Length 1081;
Best Local Similarity 87.0%; Pred. No. 1.3e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CACCTTCACAAACAAATCAATC 23
|||||
Db 613 CACCTTCACAAATCAATC 591

RESULT 15

US-10-424-599-10031
; Sequence 10031, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 10031
; LENGTH: 1132
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_109068C.1

US-10-424-599-10031

Query Match 79.1%; Score 18.2; DB 13; Length 1132;
Best Local Similarity 87.0%; Pred. No. 1.3e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CACCTTCACAAACAAATCAATC 23
|||||
Db 187 CACCTTCACAAACAAATCAATC 209

Search completed: April 24, 2004, 23:47:31
Job time : 187.753 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2004, 18:23:30 ; Search time 1790.31 Seconds

(without alignments)
383.638 Million cell updates/sec

Title: US-10-084-555A-116

Perfect score: 23
Sequence: 1 caccctcacaaaaaatcaatc 23

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estnu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_esti:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_est3:*
- 12: gb_est4:*
- 13: gb_est5:*
- 14: gb_est6:*
- 15: em_estfun:*
- 16: em_estom:*
- 17: em_gss_hum:*
- 18: em_gss_inv:*
- 19: em_gss_pin:*
- 20: em_gss_vrt:*
- 21: em_gss_fun:*
- 22: em_gss_mam:*
- 23: em_gss_mus:*
- 24: em_gss_pro:*
- 25: em_gss_red:*
- 26: em_gss_pug:*
- 27: em_gss_vri:*
- 28: gb_gss1:*
- 29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	20.4	88.7	109	12 BG359596	BG359596 sac26b07.
C 2	19.8	86.1	220	9 AV347217	AV347217 AV347217
C 3	19.8	86.1	474	13 BU545164	BU545164 GM880004A
C 4	19.8	86.1	522	14 CF442018	CF442018 EST678363

C	5	19.4	84.3	294	9	AV090955	AV090955
C	6	19.4	84.3	690	28	AZ385940	AZ385940 1M0144J08
C	7	19	82.6	422	10	BF202981	BF202981 WHE1766A
C	8	19	82.6	784	29	CG106216	CG106216 PUEK64TD
C	9	19	82.6	910	29	CG459064	CG459064 PUEK64TD
C	10	19	82.6	951	28	CC397488	CC397488 PUEK64TD
C	11	19	82.6	985	28	CNS0112H	AL100919 DROB08TB
C	12	19	82.6	1209	14	CD495335	CD495335 CDA15-H06
C	13	18.8	81.7	270	28	CC392104	CC392104 PUEK64TD
C	14	18.8	81.7	366	13	BY427980	BY427980 BY427980
C	15	18.8	81.7	400	29	CE791564	CE791564 tigr-gss-
C	16	18.8	81.7	405	14	RZ5992	RZ5992 YN44810.r1
C	17	18.8	81.7	422	10	BF705554	BF705554 RH122_4_F
C	18	18.8	81.7	490	13	BU092397	BU092397 946130C10
C	19	18.8	81.7	490	14	CB816108	CB816108 3529_1_79
C	20	18.8	81.7	517	14	CB816585	CB816585 3529_1_84
C	21	18.8	81.7	530	14	CB816482	CB816482 3529_1_79
C	22	18.8	81.7	541	14	CB834661	CB834661 3529_1_38
C	23	18.8	81.7	559	12	EG895694	EG895694 359274_MA
C	24	18.8	81.7	576	14	CA283016	CA283016 SCFRLB202
C	25	18.8	81.7	603	14	CA830818	CA830818 1117011H0
C	26	18.8	81.7	605	14	CB833808	CB833808 3529_1_84
C	27	18.8	81.7	614	12	BQ025465	BQ025465 UT-1-BB1p
C	28	18.8	81.7	630	29	CE307763	CE307763 tigr-gss-
C	29	18.8	81.7	654	14	CA231587	CA231587 SCQSPJ203
C	30	18.8	81.7	668	13	BU036865	BU036865 946130C10
C	31	18.8	81.7	678	29	CE831450	CE831450 tigr-gss-
C	32	18.8	81.7	704	29	AG095974	AG095974 Pan trol
C	33	18.8	81.7	705	14	CF243640	CF243640 3530_1_22
C	34	18.8	81.7	730	28	BZ714750	BZ714750 OGBBD54TC
C	35	18.8	81.7	768	28	BZ612817	BZ612817 WHAAR25TR
C	36	18.8	81.7	825	29	CG309152	CG309152 OGLA265TV
C	37	18.8	81.7	877	13	BQ940895	BQ940895 AGENCOURT
C	38	18.8	81.7	877	29	CG309161	CG309161 OGLA265TV
C	39	18.8	81.7	878	28	AZ668717	AZ668717 ENTML14TR
C	40	18.8	81.7	986	28	CC191636	CC191636 CH261-6L1
C	41	18.8	81.7	1463	12	BM018502	BM018502 603646330
C	42	18.4	80.0	261	9	AU060588	AU060588 AU060588
C	43	18.4	80.0	277	9	AI506637	AI506637 vn49g03.x
C	44	18.4	80.0	364	29	AG264296	AG264296 Lotus cor
C	45	18.4	80.0	374	29	CE628941	CE628941 tigr-gss-

ALIGNMENTS

RESULT 1
BG359596/c
LOCUS BG359596 109 bp mRNA linear EST 28-NOV-2001
DEFINITION sac26b07.y1 Gm-cl051 Glycine max cDNA clone GENOME SYSTEMS CLONE
ID: Gm-cl051-3158 5', mRNA sequence.
ACCESSION BG359596
VERSION BG359596.1 GI:13240287
KEYWORDS EST
SOURCE Glycine max (soybean)
ORGANISM Glycine max

REFERENCE
AUTHORS
Shoemaker, R., Keim, P., Vodkin, L., Erpelidg, J., Corvett, V.,
Khan, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J.,
Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M.,
Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N.,
Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M.,
McCann, R., Waterston, R., and Wilson, R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800
Fax: 314 286 1810

Fax: 314 288 1810
Email: est@watson.wustl.edu

Email: es@watson.wustl.edu
This clone is available through: ResGen, Invitrogen Corp. 2130
South Memorial Parkway Huntsville, AL 35801 For further information
call: (800)-533-4363 or contact via email: ccu@resgen.com
High quality sequence stop: 99.

FEATURES
SOURCE

```
Location/Qualifiers
1. 109
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1051-3168"
/tissue_type="floral meristematic mRNA"
/lab_host="DH10B"
/clone_lib="Gm-c1051"
```

/clone="_lib5-Gm-Clus1"
/note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2: XhoI; The cDNA library was constructed from floral meristematic mRNA provided by Dr. Halina Knap of Clemson University. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into DH10B host cells (GibcoBRL). This library was constructed in the laboratory of Dr. Randy Shoemaker."

ORIGIN

Query Match 88.7%; Score 20.4; DB 12; Length 109;
Best Local Similarity 95.5%; Pred. No. 2.5e+04;
Matches 21; Conservative 0; Mismatches 1; Indels 0

QY 1 CACCTTCACAAAAAATCAAT 22
|||
Dδ 68 CACCTTCACAAAAAATCAAT 47

RESULT 2

AV347217/c		linear	EST 12-NOV-1999
LOCUS	220 bp	mRNA	
DEFINITION	AV347217 RIKEN full-length enriched, adult male olfactory bulb Mus musculus CDNA clone 6430S94G07 3' similar to A070981 Mus musculus brain mitochondrial carrier protein BMCPI (Bmcpl) mRNA, mRNA sequence.		

ACCESSION	VERSION	GI
AV347217	AV347217.1	6388276

VERSION
KEYWORDS
SOURCE
AV347217.1 GT:0000270
EST.
Mus musculus (house mouse)

WORKS CITED

Mus musculus (house mouse), **ORGANISM**
Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
1 (bases 1 to 220) **REFERENCE**

REFERENCE
AUTHORS

63 CATCTTCACAAAAAACCACATC 41

RESULT 3
BU545164

SNIZATO, Y., SNLUGEMOTO, I., SHIRAKI, I., SOGABE, I., SUGANAGA, I.,
SUZUKI, H., SUZUKI, H., TAKAHASHI, F., TATENO, M., TOMINAGA, N.,
TSUNODA, Y., WATAHAKI, A., WATANABE, S., YAMAMURA, T., YASUNISHI, A.,
YOKOTA, T., YOSHIDA, A., YOSHINO, M., MURAMATSU, M. and HAYASHIZAKI, Y.
RIKEN MUSEE RSTTS (Konno, H. et al. 1999)

TITLE
JOURNAL

JOURNAL COMMENT
 Submitted: 11/29/97
 Accepted: 11/29/97
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suoh-cho, Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,
URL:<http://genome.gsc.riken.go.jp/>,
Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
Matsueura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and
Hayashizaki,Y.
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase. *Proc. Natl. Acad. Sci. U.S.A.* 95 (7), 3455-3460 (1998)
Itoh,M., Kitsuai,T., Akiyama,Y., Shibata,K., Izawa,M., Kawai,J.,
Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M.,
Okazaki,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation
system. *Genome Res.* 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. *Methods Enzymol.* 303,
19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp/>) for
further details.

FEATURES

```
1. 2221 "Mus musculus"  
/organism="Mus musculus"  
/mol_type="mRNA"  
/strain="G57BL/6J"  
/db_xref="taxon:10090"  
/clone="6430594G07"  
/sex="male"  
/tissue_type="olfactory brain"  
/dev_stage="DH10B"  
/lab_host="DH10B"  
/clone_lib="RIKEN full-length enriched, adult male  
olfactory bulb"
```

oratory build' /note-site1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGATCCAGACGCTCTTTTCTTTTCTTWN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGAGATCTCGAGTTAATAATATATCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5', end: SalI; 3', end: BamHI".

ORIGIN

Query Match	86.1%;	Score 19.8;	DB 9;	Length 220;
Best Local Similarity	91.3%;	Pred. No. 2.7e+04;		
Matches 21: Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;

Qy 1 CACCTTCACAAAAAATCAATC 23
Db 63 CATCTTCACAAAAAACCATC 41

RESULT 3	EST 16-SEP-2002
BU545164	
LOCUS	
DEFINITION	linear mRNA
	474 bp
	Glycine max cDNA clone Gm-r1088-1362 3',
	BU545164
	Gm-r1088
	max cDNA
	clone Gm-r1088-1362 3',
	DEFINITION
	mRNA sequence.

ACCESSION	BU545164
VERSION	BU545164.1
	GI:22928025

KEYWORDS EST. Glycine max (soybean)
SOURCE

ORGANISM

Glycine max

Eukaryota, Viridiplantae, Streptophyta, Embryophyta, Tracheophyta, Spermatophyta, Magnoliophyta, eudicotyledons, core eudicots, rosids, eurosids I, Fabales, Fabaceae, Papilionoideae, Phaseoleae, Glycine.

```

REFERENCE
AUTHORS
Vodkin,L., Shoemaker,R., Keim,P., Retzel,E., Khanna,A., Shealy,R.,
Clough,S., Thibaud-Nissen,F., Corryell,V., Erpelting,J., Rapp,C.,
Shoop,E., Stronvik,M., Schweitzer,P., Gong,G. and Liu,L.
A Functional Genomics Program for Soybean (NSF 9872565) (2002)
JOURNAL
Unpublished (2002)
COMMENT
Other ESTs: AW703762 corresponding to Gm-cl023-3848 (5')
Contact: Vodkin, L.O., PI, A Functional Genomics Program for
Soybean (NSF 9872565),
Lewin, H. A., Director, Keck Center for Comparative and Functional
Genomics
University of Illinois
Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
Tel: (217) 244-6147
Fax: (217) 333-4582
Email: l-vodkin@uiuc.edu
Insert length: 474 Std Error: 0.00
Plate: GM880004A20 row: E column: 09
Seq primer: 5'-TTTTTTTTTTTTTTTTTT(A/C/G)-3'
High quality sequence stop: 474.
FEATURES
Location/Qualifiers
1..474
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/db_xref="Gm-cl088-1362"
/clone="Gm-cl088"
/notes="The library Gm-cl088 is a sequence-driven, reracked
set of 9,216 low redundancy clones selected from cDNA
libraries from various tissues and stages of development
of soybean that consists of 2,706 cDNAs from germinating
cotyledons (source library Gm-cl027); 1,355 cDNAs from
immature seed coats (libraries Gm-cl019 and Gm-cl023); 917
cDNAs from tissue culture derived somatic embryos (source
libraries Gm-cl036 and Gm-cl075); 3,275 cDNAs from
germinating seedlings, shoot tips, or leaves exposed to
various stresses (source libraries Gm-cl064, Gm-cl065,
Gm-cl066, and Gm-cl067); and Gm-cl068); and 963 cDNAs from
young leaves exposed to bacterial and fungal pathogens
(source libraries Gm-cl072, Gm-cl073, and Gm-cl074). The
5' ESTs of the source clones from the different libraries
was used to select singletons, or a representative of each
contig, which were reracked to form library Gm-cl088 and
the cDNA clones of the reracked Gm-cl088 library were then
sequenced at the 3' end. The unigene selection and 3'
sequencing was funded by NSF Plant Genome project #9872565
(http://soybeanomics.cropsci.uiuc.edu/) as part of
creation of a low redundancy soybean cDNA set. The source
cDNA libraries were constructed by the laboratories of
Lila Vodkin, University of Illinois, Randy Shoemaker, Iowa
State University, and Paul Keim, Northern Arizona
University as part of the Public EST project,
http://129.186.26.94/soybeanest.html. The contig analysis
to select unique genes was performed by the laboratory of
Ernest Retzel, Center for Computational Genomics and
Bioinformatics, University of Minnesota,
http://web.ahc.umn.edu/biodata/nfssoy/. Reracking and 3'
sequencing were conducted by services of the University of
Illinois Keck Center for Comparative and Functional
Genomics http://www.biotech.uiuc.edu/keck.htm. Note: The
corresponding 5' EST from each clone in the Gm-cl088
library is listed in the 'OTHER EST' field. The detailed
information on the source library for each clone can also
be obtained by referring to the Genome Systems clone ID of
the original cDNA library that is also listed under 'OTHER
EST'."
ORIGIN
Query Match 86.1%; Score 19.8; DB 13; Length 474;
Best Local Similarity 91.3%; Pred. No. 2e+04;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 CACCTTCACAAAAAATCAATC 23

```

```

Db 234 CACCTTCACAAAAAATCAATC 256
||||| ||||||| ||||||| ||||||| |||||||
RESULT 4
LOCUS
CF442018 522 bp mRNA linear EST 04-SEP-2003
DEFINITION
ES678363 normalized cDNA library of onion Allium cepa cDNA clone
ACAPP27, mRNA sequence.
ACCESSION
CF442018
VERSION
CF442018.1 GI:34464720
KEYWORDS
EST.
SOURCE
Allium cepa (onion)
ORGANISM
Allium cepa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Alliaceae;
Allium.
REFERENCE
1 (bases 1 to 522)
HAYVEY,M.J., Cheung,F., Van Aken,S., Utterback,T. and Town,C.D.
Expressed Sequence Tags from a normalized library of mixed onion
tissues (Allium cepa)
Unpublished (2003)
JOURNAL
COMMENT
Contact: Hayve M.J.
Department of Horticulture
USDA-ARS and University of Wisconsin
1575 Linden Drive, Madison, WI 53706, USA
Tel: 608-262-1830
Fax: 608-262-4743
Email: mjhayve@facstaff.wisc.edu
TIGR sequence name ACAPP27TR. For more information:
http://hayve1ab hort.wisc.edu
Seq primer: CAG GAA ACA GCT ATG ACC.
FEATURES
Location/Qualifiers
1..522
/organism="Allium cepa"
/mol_type="mRNA"
/cultivar="Red Creole (bulbs), unknown(callus), Ebano &
Texas Legend (roots)"
/db_xref="taxon:4679"
/clone="ACAPP27"
/tissue_type="Callus, roots, and young bulbs"
/clone_lib="normalized cDNA library of onion"
/notes="Vector: pCMVSPORT6.1-cdb (Invitrogen); Site 1:
EcoRV (5'); Site 2: NotI (3'); Equal molar amounts of mRNA
from callus, roots, and young bulbs were combined to
synthesize the library. Normalization to enrich for
low-copy transcripts was performed by proprietary
techniques of Invitrogen."
ORIGIN
Query Match 86.1%; Score 19.8; DB 14; Length 522;
Best Local Similarity 91.3%; Pred. No. 1.9e+04;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 CACCTTCACAAAAAATCAATC 23
Db 348 CGCCTTCACAAAAAATCAATC 326
||||| ||||||| ||||||| ||||||| |||||||
RESULT 5
AV090955/c
LOCUS
AV090955 Mus musculus tongue C57BL/6J adult Mus musculus cDNA clone
DEFINITION
2310058B16, mRNA sequence.
ACCESSION
AV090955
VERSION
AV090955.1 GI:5238503
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 294)
Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,

```


quality sequence with phred score less than 20

Seq primer: Stratagene SK primer.

FEATURES

Location/Qualifiers
1..422
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE1766.A01.B02"
/tissue_type="Spike before anthesis"
/dev_stage="Adult plant"
/lab_host="E. coli SOLR"
/clone_lib="Wheat pre-anthesis spike cDNA library"
/note="vector: Lambda Uni-ZAP XR, excised phagemid;
Site 1: EcoRI; Site 2: XhoI; Plants were grown in the
greenhouse. Whole spike with awns trimmed, white, green
and yellow anther were collected and total RNA, and
poly(A) RNA were prepared, a cDNA library was made, and
the cDNA clones were in vivo excised to give pBluescript
phagemids in the TJ Clouse lab (Choi, Clouse, Fenton) at
the University of California, Riverside. Plasmid DNA
preparations and DNA sequencing were performed in the OD
Anderson lab (all other authors)."

ORIGIN

Query Match 82.6%; Score 19; DB 10; Length 422;
Best Local Similarity 100.0%; Pred. No. 3.5e+04;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ACCTTCACAAAAAATCA 20
|||||
DB 206 ACCTTACAAAAAATCA 188
|||||

RESULT 8

CG106216

LOCUS

DEFINITION PUBDK64TD ZM 0.6 1.0 KB Zea mays genomic clone ZMMB7a211K07,
784 bp DNA linear GSS 20-AUG-2003

CG106216

ACCESSION

KEYWORDS

SOURCE

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE

AUTHORS

Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and
Bennetzen, J.

TITLE

JOURNAL

COMMENT

Other GSSs: PUBDK64TB

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TF

Class: sheared ends.

FEATURES

Location/Qualifiers

1..784

/organism="Zea mays"

/mol_type="genomic DNA"

/strain="B73"

/db_xref="taxon:4577"

/clone_lib="ZMMB7a211K07"

/clone_lib="ZM_0.6 1.0 KB"

/note="vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high

Cot selected genomic DNA library"

ORIGIN

Query Match

Best Local Similarity 100.0%;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

2 ACCTTCACAAAAAATCA 20

DB

42 ACCTTACAAAAAATCA 60

RESULT 9

CG459064/c

LOCUS

DEFINITION

PUIKE18TDB ZM 0.6 1.0 KB Zea mays genomic clone ZMMB7a0602C11,
910 bp DNA linear GSS 17-SEP-2003

Genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE

AUTHORS

Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and
Bennetzen, J.

TITLE

JOURNAL

COMMENT

Other GSSs: PUIKE18TEB

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TR

Class: sheared ends.

FEATURES

Location/Qualifiers

1..910

/organism="Zea mays"

/mol_type="genomic DNA"

/strain="B73"

/db_xref="taxon:4577"

/clone_lib="ZMMB7a0602C11"

/clone_lib="ZM 0.6 1.0 KB"

/note="vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high

Cot selected genomic DNA library"

ORIGIN

Query Match

Best Local Similarity 100.0%;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

2 ACCTTCACAAAAAATCA 20

DB

149 ACCTTACAAAAAATCA 131

RESULT 10

CC397488/c

LOCUS

DEFINITION

PUIPH88TB ZM 0.6 1.0 KB Zea mays genomic clone ZMMB7a503008,
951 bp DNA linear GSS 19-MAY-2003

Genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE

AUTHORS

1 (bases 1 to 951)

AUTHORS Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Bennett, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and
Maize Genomics Consortium
Unpublished (2003)
Other_GSSs: FURPH88TD
Contact: Cathy Whitelaw
TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TR

Class: sheared ends.

Location/Qualifiers

1. .951

/organism="Zea mays"

/mol_type="genomic DNA"

/strain="B73"

/db_xref="taxon:4577"

/clone="ZMBR503008"

/note="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high"

Cot selected genomic DNA library"

ORIGIN

Query Match 82.6%; Score 19; DB 28; Length 951;
Best Local Similarity 100.0%; Pred. No. 2.5e+04;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 ACCTTCACAAAAAATCA 20

Db 66 ACCTTCACAAAAAATCA 48

RESULT 11

CNS011ZH/c

LOCUS

DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC
BACN07P04 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.

ACCESSION AL100919.1 GI:5612530

VERSION GSS.

KEYWORDS Drosophila melanogaster (fruit fly)

SOURCE Drosophila melanogaster

ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 985)

Genoscope.

AUTHORS Direct Submission

TITLE Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :

JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

COMMENT Determination of this BAC-end sequence was carried out as part of a

collaboration with the European Drosophila Genome Project (EDGP) -

http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC

library (Dros BAC) was made by Alain Billaud at CEPH (Centre

d'Etude du Polymorphisme Humain) with funding provided by a MRC

project grant. The DNA was prepared from embryos by Alain Bucheton

and Genevieve Payan. It has been constructed in the vector

pBelobAC11.

FEATURES

source

1. .985

/organism="Drosophila melanogaster"

/mol_type="genomic DNA"

/db_xref="taxon:7227"

/clone="BACN07P04"

/clone_lib="DrosBAC"

/plasmid="pBelobAC11"

/note="end : 17"

ORIGIN

Query Match 82.6%; Score 19; DB 29; Length 985;
Best Local Similarity 60.9%; Pred. No. 2.5e+04;
Matches 14; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1 CACCTTCACAAAAAATCAATC 23

Db 973 YACWTCACAAAAAATCAATC 951

RESULT 12

CD495335/c

LOCUS

DEFINITION CDA15-H06.y1d-s SHGC-CDA Gasterosteus aculeatus cDNA clone

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

CD495335

```

RESULT 13
CC392104
LOCUS
DEFINITION PUH81TD ZM 0.6 1.0 KB Zea mays genomic clone ZM8BTa407M17,
                270 bp DNA linear GSS 19-MAY-2003
GENOMIC SURVEY SEQUENCE.
CC392104
VERSION
KEYWORDS CC392104.1 GI:30872194
SOURCE
ORGANISM Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
            1 (bases 1 to 270)
REFERENCE
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
            Bennetzen,J.
TITLE Maize Genomics Consortium
JOURNAL Unpublished (2003)
COMMENT Contact: Cathy Whitelaw
            TIGR Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source           1..270
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZM8BTa407M17"
                     /clone_lib="ZM 0.6 1.0 KB"
                     /notes="vector: PCR-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
                     Cot selected genomic DNA library"
ORIGIN
Query Match      81.7%; Score 18.8; DB 28; Length 270;
Best Local Similarity 90.9%; Pred. No. 4.8e+04;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CACCTTCACAAAATAATCAAT 22
Db 151 CACCTTCACACAAATCAAT 172

RESULT 14
BY427980/c
LOCUS
DEFINITION BY427980 RIKEN full-length enriched, 17 days pregnant adult female
                amnion Mus musculus cDNA clone I920092P05 3', mRNA sequence.
ACCESSION BY427980
VERSION BY427980.1 GI:26704505
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1 (bases 1 to 366)
REFERENCE
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
            Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
            Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
            Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
            Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Bruscia,V.,
            Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
            Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
            Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
            Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,

```

```

Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Kongsawa,A.,
Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
Saitana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,K.,
Sandelin,A., Schneider,C., Sempie,C.A., Setou,M., Shimada,K.,
Verardo,R., Wagner,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,T.,
Wells,C., Wilming,L.G., Wlstedt,C., Wang,Y., Watanabe,Y.,
Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Teshii,Y.,
Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
Rogers,J., Birney,E. and Hayashizaki,Y.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
22354683
12466851
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,
Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,
Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,
Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,
Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and
Hayashizaki,Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
FEATURES             Location/Qualifiers
     source           1..366
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="I920092P05"
                     /sex="female"
                     /tissue type="amnion"
                     /dev stage="17 days pregnant adult"
                     /clone_lib="RIKEN full-length enriched, 17 days pregnant
                     adult female amnion"
ORIGIN
Query Match      81.7%; Score 18.8; DB 13; Length 366;
Best Local Similarity 90.9%; Pred. No. 4.2e+04;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      2 ACCTTCACAAAAAATCAATC 23
|||||
DB      111 ACCTTCACAAAAAATCAATC 90
|||||

RESULT 15
CE791564/c
LOCUS   tigr-gss-dog-17000317627354 Dog Library Canis familiaris genomic,
DEFINITION
ACCESSION CE791564
VERSION   CE791564.1 GI:37132328
KEYWORDS GSS.
SOURCE   Canis familiaris (dog)
ORGANISM Canis familiaris
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE 1 (bases 1 to 400)
AUTHORS   Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
          Ruesch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
          Venter,J.C.
TITLE     The dog genome: survey sequencing and comparative analysis
JOURNAL   Science 301 (5641), 1898-1903 (2003)
MEDLINE   22875432
PUBMED    14512627
COMMENT   Contact: Kirkness EF
          The Institute for Genomic Research
          Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
          Rockville, MD 20850, USA
          Tel: 301-838-0200
          Fax: 301-838-0208
          Email: ekirknes@tigr.org
          Class: shotgun
          Location/Qualifiers
            1..400
              /organism="Canis familiaris"
              /mol_type="genomic DNA"
              /strain="Standard Poodle"
              /db_xref="taxon:9615"
              /clone_lib="Dog Library"
              /note="Site 1: BstXI; Libraries were prepared from
              peripheral blood"

ORIGIN
Query Match      81.7%; Score 18.8; DB 29; Length 400;
Best Local Similarity 90.9%; Pred. No. 4.1e+04;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 ACCTTCACAAAAAATCAATC 23
|||||
DB      188 ACATTCAAAAAACAATCAATC 167
|||||

Search completed: April 24, 2004, 21:10:50
Job time : 1794.31 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2004, 17:02:49 ; Search time 532.889 Seconds
(without alignments)
1464.047 Million cell updates/sec

Title: US-10-084-555a-117

Perfect score: 18

Sequence: 1 tgggggagttatcgagc 18

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pi.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pi.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rod.*

36: em_htg_mam.*

37: em_htg_vrt.*

38: em_sy.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	18	100.0	787	1	AB094866	AB094866 Unculture
C 2	17	94.4	113958	9	AC068858	AC068858 Homo sapi
C 3	17	94.4	167339	2	AC023340	AC023340 Homo sapi
C 4	16.4	91.1	188057	2	AC128193	AC128193 Rattus no
C 5	16.4	91.1	196519	2	AC135761	AC135761 Rattus no
C 6	16.4	91.1	214125	10	AC083948	AC083948 Mus muscu
C 7	16	88.9	998	1	AF142818	AF142818 Unculture
C 8	16	88.9	998	1	AF424413	AF424413 Unculture
C 9	16	88.9	999	1	AF424410	AF424410 Unculture
C 10	16	88.9	1425	1	UBA536836	UBA536836 Unculture
C 11	16	88.9	1477	1	AF050572	AF050572 Unculture
C 12	16	88.9	342774	2	AC116082	AC116082 Rattus no
C 13	15.4	85.6	1959	8	BT008405	BT008405 Arabidops
C 14	15.4	85.6	2119	8	AY093112	AY093112 Arabidops
C 15	15.4	85.6	6063	6	AX323780	AX323780 Sequence
C 16	15.4	85.6	10809	1	AE013753	AE013753 Versinia
C 17	15.4	85.6	13653	6	AX346237	AX346237 Sequence
C 18	15.4	85.6	68090	2	AC113956	AC113956 Homo sapi
C 19	15.4	85.6	70554	2	AC135616	AC135616 Homo sapi
C 20	15.4	85.6	92620	8	AB026636	AB026636 Arabidops
C 21	15.4	85.6	145722	2	BX324224	BX324224 Danio rer
C 22	15.4	85.6	152188	10	AC123930	AC123930 Mus muscu
C 23	15.4	85.6	165087	9	AC145772	AC145772 Pan trogl
C 24	15.4	85.6	181152	2	AC110440	AC110440 Rattus no
C 25	15.4	85.6	185528	2	AC020803	AC020803 Mus muscu
C 26	15.4	85.6	185724	2	AC131585	AC131585 Mus muscu
C 27	15.4	85.6	191668	2	BX465191	BX465191 Danio rer
C 28	15.4	85.6	192434	10	AC122305	AC122305 Mus muscu
C 29	15.4	85.6	193252	2	AC122504	AC122504 Mus muscu
C 30	15.4	85.6	194516	9	AC024681	AC024681 Homo sapi
C 31	15.4	85.6	204050	1	AJ414154	AJ414154 Versinia
C 32	15.4	85.6	240457	2	AC116503	AC116503 Mus muscu
C 33	15.4	85.6	245603	2	AC124616	AC124616 Mus muscu
C 34	15.4	85.6	247337	2	AC098501	AC098501 Rattus no
C 35	15.4	85.6	252075	2	AC095551	AC095551 Rattus no
C 36	15.4	85.6	255004	2	AC112871	AC112871 Rattus no
C 37	15.4	85.6	258015	2	AC131816	AC131816 Rattus no
C 38	15.4	85.6	260335	2	AC115140	AC115140 Rattus no
C 39	15.4	85.6	263744	2	AF401201	AF401201 Mus muscu
C 40	15.4	85.6	295708	2	AC102357	AC102357 Mus muscu
C 41	15	83.3	3337	6	BD248880	BD248880 47 Human
C 42	15	83.3	3337	6	BD276377	BD276377 143 Human
C 43	15	83.3	174240	2	AC138552	AC138552 Cercopith
C 44	15	83.3	208632	2	AC073711	AC073711 Mus muscu
C 45	15	83.3	262304	2	BX510337	BX510337 Danio rer

ALIGNMENTS

RESULT	1	AB094866	787 bp	DNA	linear	BCT 11-DEC-2003
LOCUS	AB094866	Uncultured bacterium gene for 16S rRNA, partial sequence,				
DEFINITION		clone:OKB6.6.				
ACCESSION	AB094866					
VERSION	AB094866.1	GI:28372220				
KEYWORDS		uncultured bacterium				
SOURCE		uncultured bacterium				
ORGANISM		Bacteria; environmental samples.				
REFERENCE	1	Inagaki, F., Suzuki, M., Takai, K., Oida, H., Sakamoto, T., Aoki, K.,				
AUTHORS		Nealson, K.H. and Horikoshi, K.				
TITLE		Microbial Communities Associated with Geological Horizons in				

JOURNAL Coastal Subseaflor Sediments from the Sea of Okhotsk
 PUMED Appl. Environ. Microbiol. 69 (12), 7224-7235 (2003)
 14660370

REFERENCE

1 (bases 1 to 787)

Inagaki, F.

Direct Submission

Submitted (28-OCT-2002) Fumio Inagaki, Japan Marine Science &

Technology Center, Subground Animalcule Retrieval (SUGAR) Project,

Frontier Research System for Extremophiles; Natsushima-cho 2-15,

Yokosuka 237-0061, Japan [E-mail: inagaki@amstec.go.jp,

Tel: 81-468-67-9887, Fax: 81-468-67-9715]

Location/Qualifiers

1. 787

/organism="uncultured bacterium"

/mol_type="genomic DNA"

/isolation_source="subseaflor sediments from the Sea of

Okhotsk"

/db_xref="taxon:77133"

/environmental_sample

<1..>787

/product="16S ribosomal RNA"

ORIGIN

Query Match 100.08; Score 18; DB 1; Length 787;

Best Local Similarity 100.08; Fred. NC. 43;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TGTGGGAGTTCGAGC 18

Db 699 TGTGGGAGTTCGAGC 716

AC068858

LOCUS Homo sapiens chromosome 11, clone RP11-255G5, complete sequence.

DEFINITION AC068858

ACCESSION AC068858

VERSION AC068858.10 GI:23379153

KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 113958)

Birren, B., Nussbaum, C. and Lander, E.

Unpublished

2 (bases 1 to 113958)

Birren, B., Linton, L., Nussbaum, C., Lander, E., Abraham, H., Allen, N.,

Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,

Boguslavskiy, L., Boukhgalter, B., Brown, A., Burkett, G.,

Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,

Collins, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S.,

Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,

Galagan, J., Gardyna, S., Glade, S., Goyette, M., Graham, L.,

Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,

Howard, J.C., Iliev, I., Johnson, R., Jones, C., Kaur, L., Karatas, A.,

Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J.,

Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,

McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,

Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,

Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,

Pisani, D., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,

Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,

Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,

Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,

Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,

Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (10-MAY-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

TITLE

JOURNAL

AUTHORS

REFERENCE

AUTHORS

3 (bases 1 to 113958)

Birren, B., Nussbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,

Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B.,

Camrata, J., Chang, J., Chazaro, B., Choepel, Y., Collamore, A.,

Cook, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S.,

Far, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J.,

Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,

Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,

Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,

Liu, G., MacLean, C., Macdonald, P., Major, J., Matthews, C.,

McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V.,

Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H.,

O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,

Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,

Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P.,

Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,

Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,

Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,

Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (10-SEP-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

4 (bases 1 to 113958)

Birren, B., Nussbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,

Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B.,

Camrata, J., Chang, J., Chazaro, B., Choepel, Y., Collamore, A.,

Cook, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S.,

Far, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J.,

Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,

Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,

Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,

Liu, G., MacLean, C., Macdonald, P., Major, J., Matthews, C.,

McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V.,

Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H.,

O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,

Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,

Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P.,

Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,

Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,

Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,

Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (30-SEP-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Sep 30, 2002 this sequence version replaced gi:22773350.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L8709

Center clone name: 255_G_5

Only the first 114.0 kb of this clone are being submitted.

The remainder overlaps accession number AC092733 [WICGR project

L11786].

Location/Qualifiers

1..113958

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="11"

/map="11"

/clone="RP11-255G5"

/clone_lib="RP11-11 Human Male BAC"

1..222

/rpt_family="MSTA"

repeat_region complement(223..303)

repeat_region

* arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1978: contig of 1978 bp in length
 1979 2078: gap of unknown length
 2079 9871: contig of 7793 bp in length
 9872 17873: gap of unknown length
 17874 17973: contig of 7902 bp in length
 17974 27502: gap of unknown length
 27503 27502: contig of 9529 bp in length
 27503 27602: gap of unknown length
 27603 45890: contig of 18288 bp in length
 45891 61021: gap of unknown length
 61022 61121: contig of 15031 bp in length
 61122 82780: contig of unknown length
 82781 82880: gap of unknown length
 82881 112249: contig of 29369 bp in length
 112250 112349: gap of unknown length
 112350 167339: contig of 54990 bp in length.

FEATURES

source

1. 167339
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="11"
 /clone="RP11-12C11"

misc_feature

1. 1978
 /notes="assembly_name:Contig4"

misc_feature

2079. 9871
 /notes="assembly_name:Contig5"

misc_feature

clone_end:77
 vector_side:right

misc_feature

9972. 17873
 /notes="assembly_name:Contig6"

misc_feature

17974. 27502
 /notes="assembly_name:Contig7"

misc_feature

27603. 45890
 /notes="assembly_name:Contig8"

misc_feature

45991. 61021
 /notes="assembly_name:Contig9"

misc_feature

61122. 82780
 /notes="assembly_name:Contig10"

misc_feature

82881. 112249
 /notes="assembly_name:Contig11"

misc_feature

112350. 167339
 /notes="assembly_name:Contig12"

misc_feature

clone_end:SP6
 vector_side:left

ORIGIN

Query Match 94.4%; Score 17; DB 2; Length 167339;

Best Local Similarity 100.0%; Pred. No. 98;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGTGGGGAGTTATCGAG 17

Db 24237 TGTGGGGAGTTATCGAG 24221

RESULT 4

AC128193/c

LOCUS AC128193 188057 bp DNA linear HTG 20-NOV-2002
 DEFINITION Rattus norvegicus clone CH230-443G9, WORKING DRAFT SEQUENCE, 4
 uncloned pieces.

ACCESSION

AC128193

VERSION AC128193.3 GI:25139247

KEYWORDS HTG; HTGS; HTGS DRAFT; HTGS_FULLTOP.

SOURCE

Rattus norvegicus (Norway rat)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 188057)
 Muzny,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
 Allen,C., Allen,H., Aisbrooks,S., Amin,A., Anguiano,D.,
 Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
 Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
 Bismwal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
 Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
 Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
 Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
 Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
 Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
 Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
 Draper,H., Pagan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
 Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
 Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
 Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
 Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W.,
 Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
 Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
 Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
 Hollins,B., Howell,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
 Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
 Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
 Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
 Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
 Lorensuhwa,L., Louised,H., Lozano,R.J., Lu,X., Ma,J.,
 Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
 Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
 Mahoney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
 Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
 Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
 Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
 Nwaelelehu,O., Okwuonu,G., Olarunsegun,A., Pal,S., Parks,K.,
 Pasternak,S., Paul,H., Perez,A., Perez,L., Frankoch,C.,
 Plummer,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
 Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
 Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
 Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
 Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
 Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
 Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
 Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
 Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
 Valas,R., Vera,V., Villaseana,D., Waldron,L., Walker,B., Wang,J.,
 Wang,O., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
 Williams,G., Willson,R., Wleczkyk,R., Woodden,H., Worley,K.,
 Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
 Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
 Niederhausen,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
 Weinstein,G. and Gibbs,R.A.

TITLE

JOURNAL

REFERENCE

2 (bases 1 to 188057)

Worley,K.C.

Direct Submission

Submitted (19-JUL-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 188057)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (20-NOV-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Nov 20, 2002 this sequence version replaced gi:23195744.

The sequence in this assembly is a combination of BAC based reads

and whole genome shotgun sequencing reads assembled using Atlas

(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described

in the feature table below represents a scaffold in the Atlas

assembly (a 'contig-scaffold'). Within each contig-scaffold,

individual sequence contigs are ordered and oriented, and separated

by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GZNB
Center clone name: CH230-443G9
----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 173517 bases at least Q40

Consensus quality: 175122 bases at least Q30

Consensus quality: 176116 bases at least Q20

Estimated insert size: 176566; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 4 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 14817: contig of 14817 bp in length

* 14818 14917: gap of unknown length

* 14918 184572: contig of 169655 bp in length

* 184573 184672: gap of unknown length

* 184673 185958: contig of 1284 bp in length

* 185957 186056: gap of unknown length

* 186057 188057: contig of 2001 bp in length.

* Location/Qualifiers

FEATURES

source

1. 188057

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-443G9"

5652..6571

/note="clone boundary"

clone_end:T7

site:

end sequence:BZ175079"

13705..14817

/note="wgs_contig"

14918..17602

/note="wgs_contig"

ORIGIN

Query Match 91.18; Score 16.4; DB 2; Length 188057;

Best Local Similarity 94.44; Pred. No. 2.1e+02;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 TGTCGGGGAGTTATCGAGC 18

|||||

Db 130681 TGTCGGGGAGTTATCGAGC 130664

|||||

RESULT 5

AC135761/c

LOCUS

DEFINITION Rattus norvegicus clone CH230-34415, WORKING DRAFT SEQUENCE, 3

unordered pieces.

AC135761

AC135761.2 GI:25139094

HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE

ORGANISM

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus

1 (bases 1 to 196519)

Muzny,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,

Allen,C., Allen,H., Albrooks,S., Amin,A., Anguiano,D.,

Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,

Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,

Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,

Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,B.,

Cardenas,V., Carter,K., Cavazos,I., Cesar,H., Center,A.,

Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,

Cleaveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,

Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,

Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,

Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,

Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,

Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,

Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,

Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,

Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,

Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,

Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,

Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,

Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,

Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,

Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,

Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,

Lorensuwa,L., Loulseghe,H., Lozada,R.J., Lu,X., Ma,J.,

Maheshwari,M., Mahindartine,M., Mahmoud,M., Malloy,K., Mangum,A.,

Mangum,B., Mapa,P., Martin,K., Martin,R., Martiney,E.,

Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,

Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,

Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,

Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,

Nwackelemech,O., Okeke,G., Olarnpunsagoon,A., Pal,S., Parks,K.,

Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,

Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,

Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,

Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,

Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,

Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,

Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,

Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,

Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,

Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,

Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,

Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,

Williams,G., Willson,R., Wlezyk,R., Wooden,H., Worley,K.,

Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,

Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von

Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,

Weinstock,G. and Gibbs,R.A.

Direct Submission

Unpublished

2 (bases 1 to 196519)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (22-OCT-2002)

Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 196519)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (20-NOV-2002)

Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Nov 20, 2002 this sequence version replaced gi:24211060.

The sequence in this assembly is a combination of BAC based reads

and whole genome shotgun sequencing reads assembled using Atlas

(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described

in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

```

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
-----
Center project name: KBVC
Center clone name: CH230-34415
----- Summary Statistics
Consensus quality: 175620 bases at least Q40
Consensus quality: 177139 bases at least Q30
Consensus quality: 178188 bases at least Q20
Estimated insert size: 176674; sum-of-contigs estimation
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 3848: contig of 3848 bp in length
* 3849 3948: gap of unknown length
* 3949 195117: contig of 191169 bp in length
* 195118 195217: gap of unknown length
* 195218 196519: contig of 1302 bp in length.

```

FEATURES

```

source
1..196519
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-34415"
1..1252
/notes="wgs end extension
clone_end:Sp6"
misc_feature
3949..4946
/notes="wgs end extension
clone_end:Sp6"
misc_feature
182105..182239
/notes="clone boundary
clone_end:Sp6
site:
end sequence:BZ280097"
misc_feature
complement(190044..190326)
/notes="clone boundary
clone_end:T7
site:
end_sequence:BZ280096"

```

ORIGIN

```

Query Match          91.1%; Score 16.4; DB 2; Length 196519;
Best Local Similarity 94.4%; Pred. No. 2.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGTGGGAGTTTATCGAGC 18
    |||||
Db 47255 TGTGGGAGTTTCTCGAGC 47238

```

RESULT 6

```

AC083948      214125 bp      DNA      linear      ROD 12-OCT-2002
LOCUS      Mus musculus chromosome 5 clone RP23-67P7 strain C57BL6/J, complete
DEFINITION
AC083948
ACCESSION
VERSION
KEYWORDS
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

REFERENCE

```

AUTHORS
1 (bases 1 to 214125)
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Hahnigai,P.,
Ho,S.-L., Idol,J.R., Karlins,E., Laric,P., Lee-Lin,S.-Q.,
Legaspi,R., Maduro,Q.L., Maduro,V.B., Masiello,C., Mastrian,S.D.,
McCloskey,J.C., McDowell,J., Pearson,R., Prasad,A., Shevchenko,Y.,
Stantipop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W.,
Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Zhang,L.H. and Green,E.D.

```

TITLE

JOURNAL NISC Comparative Sequencing Initiative

REFERENCE

```

2 (bases 1 to 214125)
Green,E.D.
Direct Submission
Submitted (07-OCT-2000) NIH Intramural Sequencing Center, 8717
Grovenmont Circle, Gaithersburg, MD 20877, USA
3 (bases 1 to 214125)
Green,E.D.
Direct Submission
Submitted (05-OCT-2001) NIH Intramural Sequencing Center, 8717
Grovenmont Circle, Gaithersburg, MD 20877, USA
4 (bases 1 to 214125)
Green,E.D.
Direct Submission
Submitted (12-OCT-2002) NIH Intramural Sequencing Center, 8717
Grovenmont Circle, Gaithersburg, MD 20877, USA
On Oct 5, 2001 this sequence version replaced gi:12656793.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_zoo@nhgri.nih.gov
----- Project Information
Center project name: rr
Center clone name: 067P07

```

COMMENT

This sequence was finished as follows unless otherwise noted:
all regions were double-stranded, sequenced with an
alternate chemistry, or covered by high quality data
(i.e., phred quality >= 30); an attempt was made to resolve
all sequencing problems, such as compressions and repeats;
all regions were covered by at least one plasmid subclone
or more than one M13 subclone; and the assembly was confirmed
by restriction digest.

CLONE LENGTH: This sequence represents the entire insert of
this clone unless otherwise noted. If there are overlapping
clones, the overlaps are noted in the beginning and end of
the Features section.
Location/Qualifiers
1..214125
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL6/J"
/db_xref="taxon:10090"
/chromosome="5"
/clone="RP23-67P7"
/clone_lib="RP23"
1..3973
/note="clone overlaps with GenBank Accession Number

FEATURES

source

misc_feature

```

AC083858 (nucleotides 70746-74718) clone RP23-423A22
(center project name rg); this annotated segment
represents overlap with the finished portion of
RP23-423A22 presented in AC083858; overlap with the
unfinished portion of RP23-423A22 extends to nucleotide
45970"
misc_feature      43026..49141
                  /note="single clone coverage"
misc_feature      54707..54918
                  /note="single clone coverage"
misc_feature      183976..184164
                  /note="single clone coverage"

ORIGIN
Query Match      91.1%; Score 16.4; DB 10; Length 214125;
Best Local Similarity 94.4%; Pred. No. 2.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGTGGGGAGTTATCGAGC 18
    |||||
Db 89871 TGTGGGGAGTTATCAAGC 89888

RESULT 7
AF142818
LOCUS      Uncultured bacterium ACE-43 998 bp DNA linear BCT 10-JUL-2001
DEFINITION Uncultured bacterium ACE-43 16S ribosomal RNA gene, partial
sequence.
ACCESSION  AF142818
VERSION     1 GI:5359800
KEYWORDS    uncultured bacterium ACE-43
SOURCE      uncultured bacterium ACE-43
ORGANISM    Bacteria; Actinobacteria; environmental samples.
REFERENCE   1 (bases 1 to 998)
AUTHORS     Bowman,J.P., Rea,S.M., McCammon,S.A. and McMeekin,T.A.
TITLE       Diversity and community structure within anoxic sediment from
            marine salinity meromictic lakes and a coastal meromictic marine
            basin, Vestfold Hills, Eastern Antarctica
JOURNAL     Environ. Microbiol. 2 (2), 227-237 (2000)
MEDLINE     21114728
PUBMED      11220308
REFERENCE   2 (bases 1 to 998)
AUTHORS     Bowman,J.P., McCammon,S.A. and McMeekin,T.A.
TITLE       Direct Submission
JOURNAL     Submitted (13-APR-1999) School of Agricultural Science, University
            of Tasmania, GPO Box 252-54, Hobart, Tasmania 7001, Australia
FEATURES    source
            1..998
            /organism="uncultured bacterium ACE-43"
            /mol_type="genomic DNA"
            /db_xref="taxon:97806"
            /clone="ACE-43"
            <1..>998
            /product="16S ribosomal RNA"

ORIGIN
Query Match      88.9%; Score 16; DB 1; Length 998;
Best Local Similarity 100.0%; Pred. No. 5.9e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGGGGAGTTATCGA 16
    |||||
Db 313 TGTGGGGAGTTATCGA 328

RESULT 8
AF424413
LOCUS      Uncultured bacterium MERTZ_2CM_212 998 bp DNA linear BCT 08-MAY-2003
DEFINITION Uncultured bacterium MERTZ_2CM_212 16S ribosomal RNA gene, partial
sequence.
ACCESSION  AF424413
VERSION     1 GI:16518257
KEYWORDS    uncultured bacterium
SOURCE      uncultured bacterium
ORGANISM    Bacteria; environmental samples.
REFERENCE   1 (bases 1 to 999)
AUTHORS     Bowman,J.P. and McCuaig,R.D.
TITLE       Biodiversity, Community Structural Shifts, and Biogeography of
            Prokaryotes within Antarctic Continental Shelf Sediment
JOURNAL     Appl. Environ. Microbiol. 69 (5), 2463-2483 (2003)
MEDLINE     12732511
PUBMED      12732511
REFERENCE   2 (bases 1 to 999)
AUTHORS     Bowman,J.P., McCammon,S.A., McCuaig,R.D., Gibson,J.A.E. and
            Nichols,P.D.
TITLE       Characterization of continental shelf sediments collected off
            Antarctica: microbial metabolic activity, community structure and
            biogeography
JOURNAL     Unpublished
REFERENCE   3 (bases 1 to 999)
AUTHORS     Bowman,J.P., McCammon,S.A., McCuaig,R.D., Gibson,J.A.E. and
            Nichols,P.D.
TITLE       Direct Submission

```

```

KEYWORDS    uncultured bacterium
SOURCE      uncultured bacterium
ORGANISM    Bacteria; environmental samples.
REFERENCE   1 (bases 1 to 998)
AUTHORS     Bowman,J.P. and McCuaig,R.D.
TITLE       Biodiversity, Community Structural Shifts, and Biogeography of
            Prokaryotes within Antarctic Continental Shelf Sediment
JOURNAL     Appl. Environ. Microbiol. 69 (5), 2463-2483 (2003)
MEDLINE     12732511
PUBMED      12732511
REFERENCE   2 (bases 1 to 998)
AUTHORS     Bowman,J.P., McCammon,S.A., McCuaig,R.D., Gibson,J.A.E. and
            Nichols,P.D.
TITLE       Characterization of continental shelf sediments collected off
            Antarctica: microbial metabolic activity, community structure and
            biogeography
JOURNAL     Unpublished
REFERENCE   3 (bases 1 to 998)
AUTHORS     Bowman,J.P., McCammon,S.A., McCuaig,R.D., Gibson,J.A.E. and
            Nichols,P.D.
TITLE       Direct Submission
JOURNAL     Submitted (26-SEP-2001) School of Agricultural Science, University
            of Tasmania, GPO Box 252-54, Hobart, Tasmania 7001, Australia
FEATURES    source
            1..998
            /organism="uncultured bacterium"
            /mol_type="genomic DNA"
            /isolate="360"
            /db_xref="taxon:77133"
            /clone="MERTZ_2CM_212"
            <1..>998
            /product="16S ribosomal RNA"

ORIGIN
Query Match      88.9%; Score 16; DB 1; Length 998;
Best Local Similarity 100.0%; Pred. No. 5.9e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGGGGAGTTATCGA 16
    |||||
Db 313 TGTGGGGAGTTATCGA 328

RESULT 9
AF424410
LOCUS      Uncultured bacterium MERTZ_2CM_64 999 bp DNA linear BCT 08-MAY-2003
DEFINITION Uncultured bacterium MERTZ_2CM_64 16S ribosomal RNA gene, partial
sequence.
ACCESSION  AF424410
VERSION     1 GI:16518254
KEYWORDS    uncultured bacterium
SOURCE      uncultured bacterium
ORGANISM    Bacteria; environmental samples.
REFERENCE   1 (bases 1 to 999)
AUTHORS     Bowman,J.P. and McCuaig,R.D.
TITLE       Biodiversity, Community Structural Shifts, and Biogeography of
            Prokaryotes within Antarctic Continental Shelf Sediment
JOURNAL     Appl. Environ. Microbiol. 69 (5), 2463-2483 (2003)
MEDLINE     12732511
PUBMED      12732511
REFERENCE   2 (bases 1 to 999)
AUTHORS     Bowman,J.P., McCammon,S.A., McCuaig,R.D., Gibson,J.A.E. and
            Nichols,P.D.
TITLE       Characterization of continental shelf sediments collected off
            Antarctica: microbial metabolic activity, community structure and
            biogeography
JOURNAL     Unpublished
REFERENCE   3 (bases 1 to 999)
AUTHORS     Bowman,J.P., McCammon,S.A., McCuaig,R.D., Gibson,J.A.E. and
            Nichols,P.D.
TITLE       Direct Submission

```

```

JOURNAL Submitted (26-SEP-2001) School of Agricultural Science, University
of Tasmania, GPO Box 252-54, Hobart, Tasmania 7001, Australia
FEATURES
  source
    1. 999
      /organism="uncultured bacterium"
      /mol_type="genomic DNA"
      /isolate="357"
      /db_xref="taxon:77133"
      /clone="WERTZ_2CM_64"
      <1..>999
      /product="16S ribosomal RNA"

rRNA

ORIGIN
  Query Match      88.9%; Score 16; DB 1; Length 999;
  Best Local Similarity 100.0%; Pred. No. 5.9e+02;
  Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGTGGGGAGTTATCGA 16
    |||||
Db 313 TGTGGGGAGTTATCGA 328

RESULT 10
UBAS36936
LOCUS UBAS36936 1425 bp DNA linear BCT 16-OCT-2003
DEFINITION Uncultured bacterium partial 16S rRNA gene, clone 122.
ACCESSION AJ536936
VERSION AJ536936.1 GI:37699822
KEYWORDS 16S ribosomal RNA; 16S rRNA gene.
SOURCE uncultured bacterium
ORGANISM Bacteria; environmental samples.
REFERENCE
  1 Tonolla, M.A., Demarta, A. and Peduzzi, R.
  Global phylogeny of the microbial populations in the anoxic
  waterlayers of the meromictic Lake Cadagno
  Unpublished
  2 (bases 1 to 1425)
Tonolla, M.A.
Direct Submission
Submitted (16-JAN-2003) Tonolla M.A., Environmental microbiology,
Cantonal Institute of Microbiology, Via Mirasole 22A, CH-6500
Bellinzona, SWITZERLAND
Location/Qualifiers
  1. 1425
    /organism="uncultured bacterium"
    /mol_type="genomic DNA"
    /isolation_source="chemocline"
    /db_xref="taxon:77133"
    /clone="122"
    /environmental sample
    /country="Switzerland:Lake Cadagno"
  1. 1425
    /gene="16S rRNA"
    <1..>1425
    /gene="16S rRNA"
    /product="16S ribosomal RNA"

gene
rRNA

ORIGIN
  Query Match      88.9%; Score 16; DB 1; Length 1425;
  Best Local Similarity 100.0%; Pred. No. 5.7e+02;
  Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGTGGGGAGTTATCGA 16
    |||||
Db 842 TGTGGGGAGTTATCGA 857

RESULT 11
AF050572
LOCUS AF050572 1477 bp DNA linear BCT 15-OCT-1998
DEFINITION Uncultured eubacterium WCHB1-81 16S ribosomal RNA gene, partial
sequence.

```

```

ACCESSION AF050572
VERSION AF050572.1 GI:2967748
KEYWORDS uncultured eubacterium WCHB1-81
SOURCE Bacteria; Actinobacteria; environmental samples.
ORGANISM
  1 (bases 1 to 1477)
  Dojka, M.A., Hugenholtz, P., Haack, S.K. and Pace, N.R.
  Microbial diversity in a hydrocarbon- and
  chlorinated-solvent-contaminated aquifer undergoing intrinsic
  bioremediation
  Appl. Environ. Microbiol. 64 (10), 3869-3877 (1998)
  98432811
  9758812
  2 (bases 1 to 1477)
  Dojka, M.A. Jr., Hugenholtz, P., Haack, S. and Pace, N.R.
  Direct Submission
  TITLE Submitted (24-FEB-1998) Plant and Microbial Biology, University of
  California, Berkeley, 111 Koshland Hall, Berkeley, CA 94720, USA
  JOURNAL Location/Qualifiers
  1. 1477
    /organism="uncultured eubacterium WCHB1-81"
    /mol_type="genomic DNA"
    /db_xref="taxon:74218"
    /clone="WCHB1-81"
    /note="isolated from a contaminated aquifer"
    <1..>1477
    /product="16S ribosomal RNA"

rRNA

ORIGIN
  Query Match      88.9%; Score 16; DB 1; Length 1477;
  Best Local Similarity 100.0%; Pred. No. 5.7e+02;
  Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGTGGGGAGTTATCGA 16
    |||||
Db 818 TGTGGGGAGTTATCGA 833

RESULT 12
AC116082
LOCUS AC116082 342774 bp DNA linear HTG 19-NOV-2002
DEFINITION Rattus norvegicus clone CH230-155L19, WORKING DRAFT SEQUENCE, 2
unordered pieces.
ACCESSION AC116082
VERSION AC116082.4 GI:24635605
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM
  Rattus norvegicus
  Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
  Rattus.
  1 (bases 1 to 342774)
  Muzny, D., Marie, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J.,
  Allen, C., Allen, H., Alshrocks, S., Amin, A., Anguiano, D.,
  Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
  Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, P.,
  Biswal, K., Blair, J., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
  Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
  Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
  Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
  Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
  Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
  Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
  Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
  Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
  Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
  Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
  Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
  Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, C.,
  Harvey, I., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
  Hennessey, R., Hines, S., Hladun, S.B., Hodgson, A., Hogues, M.,
  Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,

```

Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, X., London, P., Longacre, S., Lopez, J., Loreschnew, L., Lousseng, H., Lozada, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K., Nwaokeme, O., Okwuonu, G., Olampunsgoon, A., Pal, S., Parks, K., Pasceriak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quirroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, M., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartbeyn, A., Sisson, I., Sitter, C.D., Smaje, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G., and Gibbs, R.A.

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 341221: contig of 341221 bp in length
 341222 341321: gap of unknown length
 341322 342774: contig of 1453 bp in length.

FEATURES

source

1. 342774
 /location/Qualifiers

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-155L19"

1-1157

/note="wgs contig"

134740..136023

/note="wgs contig"

137114..138461

/note="wgs contig"

ORIGIN

Query Match 88.98; Score 16; DB 2: Length 342774;
 Best Local Similarity 100.0%; Pred. No. 3.4e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GTGGGAGTTCGAG 17

Db 213095 GTGGGAGTTCGAG 213110

RESULT 13

BT008405

LOCUS

DEFINITION

Arabidopsis thaliana At3g17040 gene, complete cds.

ACCESSION

BT008405

VERSION

BT008405.1

KEYWORDS

FLI CDNA

SOURCE

Arabidopsis thaliana

ORGANISM

Arabidopsis thaliana

REFERENCE

1 (bases 1 to 1959)

AUTHORS

Cheuk, R., Chen, H., Kim, C.J., Shinn, P., Bowser, L., Carninci, P., Dale, J.M., Hayashizaki, Y., Heuan, V.W., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Lam, B., Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J., Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A., Toriumi, M., Wong, C., Wu, H.C., Yamada, K., Yu, G., Yuan, S., Shinozaki, K., Davis, R.W., Theologis, A. and Ecker, J.R.

TITLE

Arabidopsis ORF clones

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 1959)

AUTHORS

Cheuk, R., Chen, H., Kim, C.J., Shinn, P., Bowser, L., Carninci, P., Dale, J.M., Hayashizaki, Y., Heuan, V.W., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Lam, B., Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J., Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A., Toriumi, M., Wong, C., Wu, H.C., Yamada, K., Yu, G., Yuan, S., Shinozaki, K., Davis, R.W., Theologis, A. and Ecker, J.R.

TITLE

Direct Submission

JOURNAL

Submitted (15-MAY-2003)

Salk Institute Genomic Analysis Laboratory

(SIGNAL), Plant Biology Laboratory, The Salk Institute for

Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,

USA

COMMENT

RIKEN Genomic Sciences Center (GSC) members carried out the

collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN

Arabidopsis Full-length cDNA') : Seki, M., Narusaka, M., Ishida, J.,

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: G082

Center clone name: CH230-155L19

Assembly program: Phrap; version 0.990329

Consensus quality: 198149 bases at least Q40

Consensus quality: 199953 bases at least Q30

Consensus quality: 201368 bases at least Q20

Estimated insert size: 199390; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y., and Shinozaki, K.

The Salk, Stanford, PESC (SSP) Consortium members constructed and sequenced the pUNI (ORF) clones using the RAPL cDNAs: Cheuk, R., Chen, H., Kim, C.J., Shinn, P., Bowser, L., Chan, M.M., Chang, C.M., Dale, J.M., Hsuan, V.W., Jones, T., Karlin-Neumann, G., Lam, B., Lee, J.M., Lin, J., Miranda, M., Nguyen, M., Onodera, C.S., Palm, C.J., Quach, H.L., Southwick, A., Tang, C.C., Toriumi, M., Wong, C., Wu, H.C., Yamada, K., Yu, G., Yuan, S., Davis, R.W., Theologis, A., and Ecker, J.R.

Cheuk, R. (SSP/Salk) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Ecker, J.R. (SSP/Salk) contributed equally to this work as PIs.

FEATURES

source
1. 1959
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="3"
/clone="U17978"
/ecotype="Columbia"
/note="This clone is in pUNI 51"

CDS

1. 1959
/note="unknown protein"
/codon_start=1
/product="At3g17040"
/protein_id="AAP37764.1"
/db_xref="GI:30725484"
/translation="MHFFVPKSSSSPSPANTSSFLTPQIPENLCKSPTKIHI
GTHGIGSQSLSHPTSSKNTLYAVVDRSSGVSPQESANGESNTEGVLV
RRLLENDSKESSEEGKKYPARIDAGLSNIANKMPFEPERSESSSSAAAAAARQ
ERPLAVNLDLSLYKAVLARNFRYKDAEKLKIAWPEDEGRPYVALGKILSKOSKL
AEARILYKCGSQSGENSIYWCWAVLENLGNVRARELFDATVADKKHVAWHG
WANLEIKOGNISKARNLLAKGLFCGRNEVYIQTLLAEKAGRYEQARYLFGKQATIC
NSRCSALAWAOLEIQERYPAARKLFEKAVQSPKRFAMWVGVFGVGNVGRG
RKLKIGHALNPRDVLVLSGLLEVKHSSANLARRASELDRHOPVMIAGWM
EQKGTNTTARELYQRALSIDANTESARCLQWGLVLEQAGNLSARELFRSLNIN
SQSVYTWMTWQLEEDQDTERAEIRNLVFOQTEVVDASWVTGFLDIIDPALTIV
KRLNFGQNDNNRLTTLNMRTKDSQSNQPPSSAGREDIETGSGFNLDVFLRSK
LSLDPLKLDVNLDSKRLERFTRGRINGA"

ORIGIN

Query Match 85.6%; Score 15.4; DB 8; Length 1959;
Best Local Similarity 94.1%; Pred. No. 1.2e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GTGGGGAGTTATCGAGC 18

Db 1449 GTGGGGAGTTCTCGAGC 1465

RESULT 14
AY093112
LOCUS
DEFINITION
Arabidopsis thaliana unknown protein (At3g17040) mRNA, complete cds.
AY093112 2119 bp mRNA linear PLN 21-APR-2002
AY093112.1 GI:20260425

ACCESSION
VERSION
FLI_CDNA.
KEYWORDS
Arabidopsis thaliana
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 2119)
Southwick, A., Karlin-Neumann, G., Nguyen, M., Lam, B., Miranda, M., Palm, C.J., Bowser, L., Jones, T., Banh, J., Carninci, P., Chen, H., Cheuk, R., Chung, M.K., Hayashizaki, Y., Ishida, J., Kamiya, A., Kawai, J., Kim, C., Lin, J., Liu, S.X., Narusaka, M., Pham, P.K., Sakano, H., Sakurai, T., Satou, M., Seki, M., Shinn, P., Yamada, K., Shinozaki, K., Ecker, J., Theologis, A. and Davis, R.W.

TITLE JOURNAL

Direct Submission
Submitted (27-MAR-2002) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
e-mail for correspondence: arab@sequence.stanford.edu

COMMENT

RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAPL cDNAs (RAPL cDNA: "RIKEN Arabidopsis Full-length cDNA"): Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PESC (SSP) Consortium members carried out the sequencing and annotation of the RAPL cDNAs: Southwick, A., Nguyen, M., Karlin-Neumann, G., Lam, B., Miranda, M., Palm, C.J., Bowser, L., Jones, T., Banh, J., Chen, H., Cheuk, R., Chung, M.K., Kim, C., Lin, J., Liu, S.X., Pham, P.K., Sakano, H., Shinn, P., Yamada, K., Ecker, J., Theologis, A. and Davis, R.W.

Southwick, A., (SSP/Stanford) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Davis, R.W. (SSP/Stanford) contributed equally to this work as PIs.

FEATURES

source
1. 2119
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="3"
/clone="RAPL09-27-C16 (R17978)"
/ecotype="Columbia"
/note="This clone is in pBluescript"

gene

1. 2119
/gene="At3g17040"

CDS

49 2007
/gene="At3g17040"
/codon_start=1
/product="unknown protein"
/protein_id="AA013111.1"
/db_xref="GI:20260426"
/translation="MHFFVPKSSSSPSPANTSSFLTPQIPENLCKSPTKIHI
GTHGIGSQSLSHPTSSKNTLYAVVDRSSGVSPQESANGESNTEGVLV
RRLLENDSKESSEEGKKYPARIDAGLSNIANKMPFEPERSESSSSAAAAAARQ
ERPLAVNLDLSLYKAVLARNFRYKDAEKLKIAWPEDEGRPYVALGKILSKOSKL
AEARILYKCGSQSGENSIYWCWAVLENLGNVRARELFDATVADKKHVAWHG
WANLEIKOGNISKARNLLAKGLFCGRNEVYIQTLLAEKAGRYEQARYLFGKQATIC
NSRCSALAWAOLEIQERYPAARKLFEKAVQSPKRFAMWVGVFGVGNVGRG
RKLKIGHALNPRDVLVLSGLLEVKHSSANLARRASELDRHOPVMIAGWM
EQKGTNTTARELYQRALSIDANTESARCLQWGLVLEQAGNLSARELFRSLNIN
SQSVYTWMTWQLEEDQDTERAEIRNLVFOQTEVVDASWVTGFLDIIDPALTIV
KRLNFGQNDNNRLTTLNMRTKDSQSNQPPSSAGREDIETGSGFNLDVFLRSK
LSLDPLKLDVNLDSKRLERFTRGRINGA"

ORIGIN

Query Match 85.6%; Score 15.4; DB 8; Length 2119;
Best Local Similarity 94.1%; Pred. No. 1.2e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GTGGGGAGTTATCGAGC 18

Db 1497 GTGGGGAGTTCTCGAGC 1513

RESULT 15

AX323780
LOCUS
DEFINITION
Sequence 268 from Patent WO0192565.
ACCESSION
AX323780
VERSION
AX323780.1 GI:18094528

KEYWORDS
synthetic construct

SOURCE
synthetic construct

ORGANISM
artificial sequences.

REFERENCE
1

linear PAT 07-JAN-2002

AUTHORS Olek A., Piepenbrock C. and Berlin K.
TITLE Diagnosis of diseases associated with dna transcription
JOURNAL Patent: WO 012563-A 268 06-DEC-2001;
EpiGenomics AG (DE)

FEATURES
source
1. .6063
Location/Qualifiers
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"

ORIGIN

Query Match 85.6%; Score 15.4; DB 6; Length 6063;
Best Local Similarity 94.1%; Pred. No. 1.1e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGTGGGGAGTTATCGAG 17
Db 489 TGAGGGGAGTTATCGAG 505

Search completed: April 24, 2004, 19:25:30
Job time : 536.889 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2004, 17:01:15 ; Search time 152.222 Seconds
(without alignments)
502.342 Million cell updates/sec

Title: US-10-084-555A-117

Perfect score: 18

Sequence: 1 tctggggagttatcgagc 18

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq_29Jan04:*

1: Geneseqn1980s:*

2: Geneseqn1990s:*

3: Geneseqn2000s:*

4: Geneseqn2001as:*

5: Geneseqn2001bs:*

6: Geneseqn2002s:*

7: Geneseqn2003as:*

8: Geneseqn2003bs:*

9: Geneseqn2003cs:*

10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	15.4	85.6	587	6	ABQ35536 Oligonucleotide for detecting cytosine methylation
2	15.4	85.6	587	6	ABQ35537 Oligonucleotide for detecting cytosine methylation
3	15.4	85.6	2065	3	AAC44909 Arabidopsis
4	15.4	85.6	6063	6	ABK28394 DNA trans
5	15.4	85.6	19653	6	ABL33335 Human imm
6	15	83.3	3337	3	AAV78421 Human sec
7	15	83.3	3337	3	AAC95857 Human sec
8	15	83.3	3337	7	ADA56410 Gene enco
9	15	83.3	3337	7	ADA40248 Human sec
10	15	83.3	3337	9	ADC73765 Human sec
11	14.8	82.2	468	8	ACH37009 Human end
12	14.8	82.2	482	3	AAC40987 Zea may
13	14.8	82.2	542	6	ABL37374 Human col
14	14.8	82.2	565	3	RAC52240 Arabidops
15	14.8	82.2	565	3	RAC54690 Arabidops
16	14.8	82.2	655	6	ABQ21639 Oligonucleotide for determining the degree of
17	14.8	82.2	655	6	ABQ21638 Oligonucleotide for determining the degree of
18	14.8	82.2	738	5	AAH67258 C glutami
19	14.8	82.2	861	4	AAF71943 Corynebact
20	14.8	82.2	862	5	AAV71206 DNA enco
21	14.8	82.2	1015	3	RAC59665 Human sec
22	14.8	82.2	1204	6	ABQ42502 Oligonucleotide for determining the degree of
23	14.8	82.2	1204	6	ABQ42503 Oligonucleotide for determining the degree of

24	14.8	82.2	1635	6	ABK94905 Human nov
25	14.8	82.2	1772	3	AAC75163 Human ORF
26	14.8	82.2	1932	9	ADE56603 Human gen
27	14.8	82.2	1932	9	ADE56607 Human gen
28	14.8	82.2	1932	9	ADE56611 Human gen
29	14.8	82.2	1932	9	ADE56599 Human gen
30	14.8	82.2	2193	7	ACA34445 Prokaryot
31	14.8	82.2	2325	5	AAV77863 DNA enco
32	14.8	82.2	3028	4	ABL14830 Drosophil
33	14.8	82.2	3361	4	AAF27860 Human bre
34	14.8	82.2	3361	8	ACD40263 Human bre
35	14.8	82.2	3671	4	AAV21314 Human CDN
36	14.8	82.2	3671	7	ACD23923 Novel hum
37	14.8	82.2	3671	7	ACA67064 CDNA enco
38	14.8	82.2	3671	7	ACA03673 CDNA enco
39	14.8	82.2	3671	7	ABX89211 DNA enco
40	14.8	82.2	3671	7	ACD41865 Human sec
41	14.8	82.2	3671	7	ACA04094 Human CDN
42	14.8	82.2	3671	8	ADA45660 Novel hum
43	14.8	82.2	3671	8	ADA76091 Human PRO
44	14.8	82.2	3671	8	ADA18741 Human PRO
45	14.8	82.2	3671	8	ADA61364 Homo espi

ALIGNMENTS

RESULT 1
ABQ35536
ID ABQ35536 standard; DNA; 587 BP.
XX
AC ABQ35536;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 22127.
XX
KW Human; cytosine methylation; 5'-CG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP010074.
XX
PR 01-SEP-2000; 2000DE-01043826.
XX
PA 05-SEP-2000; 2000DE-01044543.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
XX Determining the degree of cytosine methylation in genomic DNA, useful for
XX diagnosis and prognosis, comprises selective hybridization of amplicons
XX from chemically treated DNA.
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
XX This invention describes a novel method for determining the degree of
XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a
XX genomic sample of DNA. The sample is treated chemically to convert
XX cytosine (C) but not methylated C, to uracil, then part of the genomic
XX DNA that contains the target C is amplified to form a labeled amplicon.
XX The amplicon is hybridised to two classes, each with at least one member,
XX of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
XX degree of hybridisation to both classes is determined from the label on

CC the amplicon. From the ratio of labels hybridised to the two classes of
 CC oligomers, the degree of methylation is calculated. The method is used:
 CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
 CC and of a wide range of diseases, e.g. cancer, disorders of the central
 CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
 CC particularly by detecting mutations or single nucleotide polymorphisms
 CC (SNP's); and (ii) for differentiation of cell or tissue types and for
 CC investigating cell differentiation. The method allows the methylation
 CC status of many C residues to be determined simultaneously. ABQ13410-
 CC ABQ54121 represent genomic DNA sequences used to illustrate the method
 CC for determining the degree of cytosine methylation described in the
 CC disclosure of the invention

XX SQ Sequence 587 BP; 85 A; 72 C; 245 G; 185 T; 0 U; 0 Other;

Query Match 85.6%; Score 15.4; DB 6; Length 587;
 Best Local Similarity 94.1%; Pred. No. 1.2e+02;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGTGGGAGTTATCGAG 17
 |||||
 DB 237 TGTGGGAGTTATCGAG 253

RESULT 2

ABQ35537/C
 ID ABQ35537 standard; DNA; 587 BP.

XX AC ABQ35537;

XX DT 12-JUL-2002 (first entry)

XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 22128.

XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KW drug; side effect; cancer; central nervous system; cardiovascular;
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;
 KW SNP; cell differentiation; ds.

XX OS Homo sapiens.

XX FN WO200218632-A2.

XX PD 07-MAR-2002.

XX PF 01-SEP-2001; 2001WO-EP010074.

XX PR 01-SEP-2000; 2000DE-01043826.

XX PR 05-SEP-2000; 2000DE-01044543.

XX PA (EPIG-) EPIGENOMICS AG.

XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;

XX DR WPI; 2002-371829/40.

XX PT Determining the degree of cytosine methylation in genomic DNA, useful for
 PT diagnosis and prognosis, comprises selective hybridization of amplicons
 PT from chemically treated DNA.

XX PS Claim 12; 56pp + Sequence Listing; 56pp; German.

XX CC This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one member,
 CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
 CC degree of hybridisation to both classes is determined from the label on
 CC the amplicon. From the ratio of labels hybridised to the two classes of
 CC oligomers, the degree of methylation is calculated. The method is used:
 CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs

CC and of a wide range of diseases, e.g. cancer, disorders of the central
 CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
 CC particularly by detecting mutations or single nucleotide polymorphisms
 CC (SNP's); and (ii) for differentiation of cell or tissue types and for
 CC investigating cell differentiation. The method allows the methylation
 CC status of many C residues to be determined simultaneously. ABQ13410-
 CC ABQ54121 represent genomic DNA sequences used to illustrate the method
 CC for determining the degree of cytosine methylation described in the
 CC disclosure of the invention

XX SQ Sequence 587 BP; 185 A; 245 C; 72 G; 85 T; 0 U; 0 Other;

Query Match 85.6%; Score 15.4; DB 6; Length 587;
 Best Local Similarity 94.1%; Pred. No. 1.2e+02;
 Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGTGGGAGTTATCGAG 17
 |||||
 DB 351 TGTGGGAGTTATCGAG 335

RESULT 3

AAC44309
 ID AAC44909 standard; DNA; 2065 BP.

XX AC AAC44909;

XX DT 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 44587.

XX KW Hybridisation assay; genetic mapping; gene expression control;
 KW protein identification; signal transduction pathway; metabolic pathway;
 KW promoter; termination sequence; ss.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-00301439.

XX PR 25-FEB-1999; 99US-0121825P.

XX PR 05-MAR-1999; 99US-0123180P.

XX PR 09-MAR-1999; 99US-0123548P.

XX PR 23-MAR-1999; 99US-0125788P.

XX PR 25-MAR-1999; 99US-0126264P.

XX PR 29-MAR-1999; 99US-0126785P.

XX PR 01-APR-1999; 99US-0127462P.

XX PR 06-APR-1999; 99US-0128234P.

XX PR 08-APR-1999; 99US-0128714P.

XX PR 16-APR-1999; 99US-0128845P.

XX PR 19-APR-1999; 99US-0130077P.

XX PR 21-APR-1999; 99US-0130449P.

XX PR 23-APR-1999; 99US-0130510P.

XX PR 28-APR-1999; 99US-0130891P.

XX PR 28-APR-1999; 99US-0131449P.

XX PR 30-APR-1999; 99US-0132048P.

XX PR 30-APR-1999; 99US-0132407P.

XX PR 04-MAY-1999; 99US-0132484P.

XX PR 05-MAY-1999; 99US-0132485P.

XX PR 06-MAY-1999; 99US-0132486P.

XX PR 07-MAY-1999; 99US-0132487P.

XX PR 11-MAY-1999; 99US-0132566P.

XX PR 14-MAY-1999; 99US-0134218P.

XX PR 14-MAY-1999; 99US-0134219P.

XX PR 14-MAY-1999; 99US-0134221P.

XX PR 14-MAY-1999; 99US-0134370P.

XX PR 18-MAY-1999; 99US-0134768P.

XX PR 19-MAY-1999; 99US-0134941P.

XX PR 20-MAY-1999; 99US-0135124P.

PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 21-JUN-1999; 99US-0139899P.
PR 22-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 28-JUN-1999; 99US-0140891P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 08-JUL-1999; 99US-0142820P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145132P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.

PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154019P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 28-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159323P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160981P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.

[illegible]

SQ Sequence 19653 BP; 5918 A; 187 C; 4299 G; 9249 T; 0 U; 0 Other;

Query Match 85.6%; Score 15.4; DB 6; Length 19653;
Best Local Similarity 94.1%; Pred. No. 1.7e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGTGGGAGTTATCGAG 17
|||||
DB 17158 TGTGGGAGTTATAGAG 17174

RESULT 6

AA78421/c
ID AAA78421 standard; cDNA; 3337 BP.

XX AC AAA78421;

XX DT 20-NOV-2000 (first entry)

XX DE Human secreted protein gene 41 SEQ ID NO:51.

XX KW Human; secreted protein; cytostatic; antianaemic; antidiabetic;
KW antiinflammatory; ophthalmological; antirheumatic; antiarthritic;
KW antipsoriatic; antiangiogenic; cardiant; anti-HIV; neutropic;
KW neuroprotective; antimicrobial; antiparkinsonian; cancer;
KW immune system disorder; angiogenesis; hyperproliferative disorder;
KW cardiovascular disorder; apoptosis; neurological disease;
KW infectious disease; wound healing; ss.

XX OS Homo sapiens.

XX PN WO200035937-A1.

XX PD 22-JUN-2000.

XX PF 16-DEC-1999; 99WO-US029950.

XX PR 17-DEC-1998; 98US-0112809P.

XX PR 18-DEC-1998; 98US-0113006P.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Ruben SM, Ebner R, Rosen CA, Endress GA, Soppet DR, Ni J,
PI Duan DR, Moore PA, Shi Y, Larleu DW, Olsen HS, Florence K;

XX WPI; 2000-431566/37.

XX P-PSDB; AAB24477.

XX PT Forty seven human nucleic acids encoding secreted proteins, useful in the
PT treatment, prevention and diagnosis of cancers, disorders of the immune
PT system, angiogenesis disorders, neurological diseases and
PT hyperproliferative disorders.

XX PS Claim 1; Page 470-471; 562pp; English.

XX CC The polynucleotide sequence given in AAA78381 to AAA78432 encode the
CC human secreted proteins given in AAB24437 to AAB24604. Human secreted
CC proteins have activities based on the tissues and cells the genes are
CC expressed in. Examples of activities include: cytostatic; antianaemic;
CC antidiabetic; antiinflammatory; ophthalmological; antirheumatic;
CC antiarthritic; antipsoriatic; antiangiogenic; cardiant; anti-HIV;
CC neutropic; neuroprotective; antimicrobial and antiparkinsonian. Human
CC secreted protein polynucleotides, polypeptides, antagonists and/or
CC agonists may be useful in treating, preventing, and/or diagnosing other
CC diseases, disorders, and/or conditions such as: (a) cancers; (b)
CC disorders of the immune system; (c) angiogenesis disorders; (d)
CC hyperproliferative disorders; (e) cardiovascular diseases; (f) diseases
CC associated with increase apoptosis; (g) neurological diseases; and (h)
CC infectious diseases. They are also used to promote wound healing.
CC AAA78372 to AAA78380 and AAB24436 represent sequences used in the
CC exemplification of the present invention

XX SQ Sequence 3337 BP; 998 A; 741 C; 638 G; 948 T; 0 U; 12 Other;

Query Match 83.3%; Score 15; DB 3; Length 3337;

Best Local Similarity 88.2%; Pred. No. 2.4e+02;

Matches 15; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGTGGGAGTTATCGAG 17
|||||
DB 2999 TGTGGGAGTTCTCGRG 2983

RESULT 7

AAC99857/c

ID AAC99857 standard; cDNA; 3337 BP.

XX AC AAC99857;

XX DT 13-MAR-2001 (first entry)

XX DE Human secreted protein gene 40 SEQ ID NO:50.

XX KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
KW cerebroprotective; neutropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; gene therapy; pathological condition;
KW autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;
KW neoplasm; cardiovascular disorder; cardiac arrest; cerebral ischaemia;
KW cerebrovascular disorder; angiogenesis; nervous system disorder;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; skin aging; food additive; preservative; ss.

XX OS Homo sapiens.

XX PN WO200070042-A1.

XX PD 23-NOV-2000.

XX PF 11-MAY-2000; 2000WO-US012788.

XX PR 13-MAY-1999; 99US-0134068P.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Rosen CA, Ruben SM, Moore PA, Young PE, Komatsoulis GA, Birse CE;
PI Duan RD, Florence KA, Soppet DR;

XX WPI; 2000-679828/66.

XX P-PSDB; AAB56116.

XX PT Isolated nucleic acid molecule encoding a human secreted protein is used
PT in preventing, treating or ameliorating a medical condition.

XX PS Claim 1; Page 878-879; 1065pp; English.

XX CC The polynucleotide sequences given in AAC99818 to AAC99977 encode the
CC human secreted proteins given in AAB56077 to AAB56362. Human secreted
CC proteins have activities based on the tissues and cells the genes are
CC expressed in. Examples of activities include: immunosuppressive;
CC antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
CC vasotropic; cerebroprotective; neutropic; neuroprotective; antibacterial;
CC virucide; fungicide; and ophthalmological. The human secreted
CC polynucleotides and proteins can be used to prevent, treat or ameliorate
CC a medical condition in e.g. humans, mice, rabbits, goats, horses, cats,
CC dogs, chickens or sheep. They are also used in diagnosing a pathological
CC condition or susceptibility to a pathological condition. Disorders which
CC are diagnosed or treated include autoimmune diseases e.g. rheumatoid
CC arthritis, hyperproliferative disorders e.g. neoplasms of the breast or
CC liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular
CC disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders
CC e.g. Alzheimer's disease, infections caused by bacteria, viruses and
CC fungi and ocular disorders e.g. corneal infection. The proteins can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to

CC regenerate tissues and in chemoraxis. The proteins can also be used as a
 CC food additive or preservative to increase or decrease storage
 CC capabilities. AAC9909 to AAC9917 and AAB56076 represent sequences used
 CC in the exemplification of the present invention

XX SQ Sequence 3337 BP; 998 A; 741 C; 638 G; 948 T; 0 U; 12 Other;

Query Match 83.3%; Score 15; DB 3; Length 3337;
 Best Local Similarity 88.2%; Pred. No. 2.4e+02;
 Matches 15; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGTGGGGAGTTATCGAG 17
 |||||
 Db 2999 TGTGGGGAGTTCTCGRG 2983

RESULT 8
 ADA56410/c
 ID ADA56410 standard; DNA; 3337 BP.

XX AC ADA56410;

XX DT 20-NOV-2003 (first entry)

XX DE Gene encoding human secreted protein #589.

XX immunosuppressive; antiinflammatory; antiasthmatic; antiallergic;
 KW cytosolic; cerebroprotective; neuroprotective; nootropic;
 KW cardiovascular; antiarteriosclerotic; gene therapy;
 KW human secreted protein; immune disorder; inflammation;
 KW respiratory disorder; cancer; CNS disorder; neurodegenerative disorders;
 KW inflammatory bowel disease; nephritis; Crohn's disease; asthma; allergy;
 KW multiple sclerosis; ischaemic brain injury; Parkinson's disease;
 KW Alzheimer's disease; atherosclerosis; myocarditis; chromosome mapping;
 KW triple helix formation; antisense gene therapy; forensic biology; ds;
 KW gene.

XX OS Homo sapiens.

XX FN WO2002102994-A2.

XX PD 27-DEC-2002.

XX PF 19-MAR-2002; 2002WO-US008278.

XX PR 21-MAR-2001; 2001US-0277340P.

XX PR 19-JUL-2001; 2001US-0306171P.

XX PR 13-NOV-2001; 2001US-0331287F.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Rosen CA, Ruben SM;

XX DR WPI; 2003-167512/16.

XX DR P-PSDB; ADA57306.

XX New human secreted polypeptides and polynucleotides, useful for

PT diagnosing, treating or preventing e.g. immune disorders, inflammatory

PT conditions, respiratory disorders, cancers, CNS disorders, or

PT neurodegenerative disorders, or polypeptides comprising an amino acid

PT sequence at least 95% identical to the new sequences. The polypeptides,

PT antibodies or antibody fragments that bind to the polypeptides, nucleic

PT acids encoding the polypeptides, agonists or antagonists that binds to

PT the polypeptide, are useful in preparing diagnostic or pharmaceutical

PT compositions for diagnosing, treating or preventing an e.g. immune

PT disorders, inflammatory conditions (e.g. inflammatory bowel disease,

CC nephritis or Crohn's disease), respiratory disorders (e.g. asthma and
 CC allergy), cancers (e.g. gastric, ovarian or lung cancer), CNS disorders
 CC (e.g. multiple sclerosis or ischaemic brain injury), neurodegenerative
 CC disorders (e.g. Parkinson's disease or Alzheimer's disease), and
 CC cardiovascular disorders (e.g. atherosclerosis or myocarditis). The
 CC polynucleotides are useful for chromosome identification, chromosome
 CC mapping, for controlling gene expression through triple helix formation
 CC or antisense DNA or RNA, in gene therapy, for identifying individuals
 CC from minute biological samples, in forensic biology, and as hybridization
 CC probes. The polypeptides are useful for as molecular weight markers on
 CC sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)
 CC gels, to raise antibodies, for testing biological activities, and for
 CC treating or preventing neural disorders, immune system disorders,
 CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
 CC renal, proliferative and/or cancerous diseases. This sequence corresponds
 CC to a gene encoding one of the polypeptide of the invention. Note: The
 CC sequence data for this patent did form part of the printed specification,
 CC but was obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 3337 BP; 998 A; 741 C; 638 G; 948 T; 0 U; 12 Other;

Query Match 83.3%; Score 15; DB 7; Length 3337;

Best Local Similarity 88.2%; Pred. No. 2.4e+02;

Matches 15; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGTGGGGAGTTATCGAG 17

Db 2999 TGTGGGGAGTTCTCGRG 2983

RESULT 9

ADA40248/c

ID ADA40248 standard; cDNA; 3337 BP.

XX AC ADA40248;

XX DT 20-NOV-2003 (first entry)

XX DE Human secreted protein encoding cDNA.

XX Human; secreted protein; cancer; hyperproliferative disorder;

XX rheumatoid arthritis; autoimmune disorder; haematopoietic disorder;

XX anaemia; allergic reaction; asthma; cardiovascular disorder;

XX wound healing; cytostatic; immunosuppressive; nootropic; neuroprotective;

XX antiviral; antiallergic; hepatotropic; antidiabetic; antiinflammatory;

XX vulnery; cardiac; gene therapy; ss.

XX OS Homo sapiens.

XX FN WO2002102993-A2.

XX PD 27-DEC-2002.

XX PF 19-MAR-2002; 2002WO-US008123.

XX PR 21-MAR-2001; 2001US-0277340P.

XX PR 19-JUL-2001; 2001US-0306171P.

XX PR 13-NOV-2001; 2001US-0331287F.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Rosen CA, Ruben SM;

XX DR WPI; 2003-175238/17.

XX New human secreted proteins and nucleic acid molecules, useful for

PT preparing a diagnostic or pharmaceutical composition for diagnosing,

PT preventing or treating cancer or other hyperproliferative disorder,

PT asthma, allergies or AIDS.

XX Claim 9; SEQ ID NO 630; 3205pp; English.

CC The invention relates to novel genes ADA39629-ADA40565 and proteins
 CC ADA40566-ADA41501 for human secreted proteins, useful for preventing,
 CC treating or ameliorating medical conditions e.g. by protein or gene
 CC therapy. The polypeptides, nucleic acid molecules, antibodies or their
 CC fragments, and agonists or antagonists that bind to the polypeptide are
 CC useful for preparing a diagnostic or pharmaceutical composition for
 CC diagnosing or treating cancer or other hyperproliferative disorder. The
 CC polypeptides and nucleic acid molecules are also useful for detecting,
 CC preventing, diagnosing, prognosticating, treating or ameliorating cancer
 CC or other hyperproliferative disorders including neoplasms, autoimmune
 CC disorders (e.g. diabetes, rheumatoid arthritis, systemic lupus
 CC erythematosus, multiple sclerosis, autoimmune thyroiditis or haemolytic
 CC anaemia), haematopoietic or haematological disorders (e.g. anaemia,
 CC thrombocytopenia), allergic reactions including asthma or eczema,
 CC inflammatory disorders (e.g. ischaemia-reperfusion injury, inflammatory
 CC bowel disease or Crohn's disease), neurodegenerative disorders (e.g.
 CC Alzheimer's disease or Parkinson's disease), cardiovascular disorders
 CC (e.g. atherosclerosis, myocarditis), infectious diseases (bacterial,
 CC fungal or viral infections including HIV/AIDS), or wound healing and
 CC disorders of epithelial cell proliferation. The nucleic acids are also
 CC useful for chromosome identification, radiation hybrid mapping or long-
 CC range restriction mapping, as molecular weight markers, or as
 CC hybridization or diagnostic probes. The polypeptides and antibodies are
 CC useful for providing immunological probes for differential identification
 CC of the tissues immunohistochemistry assays. Note: The sequence data for
 CC this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pat_sequences.

XX Sequence 3337 BP; 998 A; 741 C; 638 G; 948 T; 0 U; 12 Other;

XX Query Match 83.3%; Score 15; DB 7; Length 3337;

XX Best Local Similarity 88.2%; Pred. No. 2.4e+02;

XX Matches 15; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGTGGGAGTTATCGAG 17

Db 2999 TGTGGGAGTTCTCGG 2983

RESULT 10

ADCT3765/c

ID ADCT3765 standard; DNA; 3337 BP.

XX AC ADCT3765;

XX 01-JAN-2004 (first entry)

XX Human secreted protein-related DNA - SEQ ID 398.

XX antianaemic; antirheumatic; antiarthritic; antiinflammatory; antithyroid;
 KW antidiabetic; immunosuppressive; dermatological; nephrotropic;
 KW antiparkinsonian; neuroprotective; nootropic; antibacterial; virucide;
 KW fungicide; antiparasitic; antiarteriosclerotic; vulnerary; cytostatic;
 KW haemopoietic; haematologic; anaemia; autoimmune disorder;
 KW rheumatoid arthritis; inflammation; Grave's disease; diabetes;
 KW systemic lupus erythematosus; glomerulonephritis; neurodegenerative;
 KW Parkinson's; Alzheimer's; wound; hyperproliferative; atherosclerosis;
 KW cancer; bacterial; viral; fungal; parasitic infection; gene therapy;
 KW human; gene; ds.

XX Homo sapiens.

XX WO2003038063-A2.

XX 08-MAY-2003.

XX 19-MAR-2002; 2002WO-US008277.

XX 21-MAR-2001; 2001US-0277340P.

XX 19-JUL-2001; 2001US-0306171P.

XX 13-NOV-2001; 2001US-0331287P.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Ruben SM;

WPI: 2003-430516/40.

P-PSDB; ADC74380.

New human secreted polypeptide for diagnosing, preventing or treating
 PT hematopoietic or hematologic disorders (e.g. anemia), autoimmune
 PT disorders (e.g. diabetes) or hyperproliferative disorders (e.g. cancer or
 PT atherosclerosis).

Claim 27; SEQ ID NO 398; 2272bp; English.

The invention relates to a novel human secreted polypeptide comprising a
 CC defined sequence given in the specification. The polypeptide, nucleic
 CC acid molecule, antibody, agonist or antagonist of the invention may be
 CC useful for preparing a composition for diagnosing or treating a
 CC haematopoietic or haematologic disorder such as anaemia, autoimmune
 CC disorders such as rheumatoid arthritis, inflammation, Grave's disease,
 CC diabetes, systemic lupus erythematosus or glomerulonephritis,
 CC neurodegenerative disorders including Parkinson's disease and Alzheimer's
 CC disease, wounds and hyperproliferative disorders including
 CC atherosclerosis or cancer, as well as bacterial, viral, fungal or
 CC parasitic infections. The polypeptide may also be used during gene
 CC therapy procedures and for identifying a binding partner by contacting
 CC the polypeptide with a binding partner and determining whether the
 CC binding partner increases or decreases the activity of the polypeptide.
 CC The current sequence is that of the human secreted protein-related DNA of
 CC the invention.

XX Sequence 3337 BP; 998 A; 741 C; 638 G; 948 T; 0 U; 12 Other;

XX Query Match 83.3%; Score 15; DB 9; Length 3337;

XX Best Local Similarity 88.2%; Pred. No. 2.4e+02;

XX Matches 15; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGTGGGAGTTATCGAG 17

Db 2999 TGTGGGAGTTCTCGG 2983

RESULT 11

ACH37009

ID ACH37009 standard; cDNA; 468 BP.

XX AC ACH37009;

XX DT 13-OCT-2003 (first entry)

XX Human endothelial cell cDNA #5142.

XX Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
 KW genome mapping; biodiversity; genetic disorder.

XX Homo sapiens.

XX US2003073623-A1.

XX 17-APR-2003.

XX 30-JUL-2001; 2001US-00918995.

XX 30-JUL-2001; 2001US-00918995.

XX (DRMA/) DRMANAC R T.

XX (LABA/) LABAT I.

XX (STAC/) STACHE-CRAIN B.

XX (DICK/) DICKSON M C.

XX (JONE/) JONES L W.

XX Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;

```

DR WPI; 2003-615964/58.
XX
PT New polynucleotide sequences obtained from various cDNA libraries, useful
PT as hybridization probes, as oligomers for PCR, for chromosome and gene
PT mapping, in the recombinant production of protein, or in generating
PT antisense DNA or RNA.
XX
PS Claim 1; SEQ ID NO 24221; 44pp; English.
XX
CC The invention relates to an isolated polynucleotide comprising any one of
CC 38043 cDNA sequences, appearing as ACH2789-ACH50831, whose sequence was
CC determined by the technique of SH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostics as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversity, or in identifying mutations
CC responsible for genetic disorders and other traits. The nucleotide
CC sequences are also useful as hybridisation probes, as oligomers for PCR,
CC for chromosome and gene mapping, in the recombinant production of
CC protein, or in generating antisense DNA or RNA. The purified polypeptide
CC is useful for generating antibodies specific for it. The present sequence
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?docid=20030073623
XX
SQ Sequence 468 BP; 111 A; 121 C; 131 G; 104 T; 0 U; 1 Other;
Query Match 82.2%; Score 14.8; DB 8; Length 468;
Best Local Similarity 88.9%; Pred. No. 2.6e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 TGTGGGAGTATCGAGC 18
DB 315 TGTGGGAGTATCAAC 332

RESULT 12
AAC40987/c
ID AAC40987 standard; DNA; 482 BP.
XX
AC AAC40987;
XX
DT 17-OCT-2000 (first entry)
XX
DE Zea mays DNA fragment SEQ ID NO: 30242.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic; pathway;
KW promoter; termination sequence; corn; ss.
XX
OS Zea mays subsp. mays.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0132863P.
PR 14-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136394P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 02-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 18-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.

```

```
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 22-JUL-1999; 99US-0145193P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149426P.
PR 18-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157863P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.

PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160982P.
PR 22-OCT-1999; 99US-0160983P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match      82.2%; Score 14.8; DB 3; Length 482;
Best Local Similarity 88.9%; Pred. No. 2.6e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGTGGGAGTTATCGAGC 18
DB 66 TGTGGGAGTTGTGGAGC 49

RESULT 13
ABL37374
ID ABL37374 standard; cDNA; 542 BP.
AC ABL37374;
XX
XX 08-APR-2002 (first entry)
XX Human colon tumour antigen polynucleotide SEQ ID NO:963.
XX Human; colon cancer; colon tumour antigen; cytostatic; vaccine;
XX colon tumour metastatic antigen; diagnosis; gene; ss.
XX Homo sapiens.
XX WC200196388-A2.
XX
XX 20-DEC-2001.
XX
XX 08-JUN-2001; 2001WO-US018557.
XX
XX 09-JUN-2000; 2000US-0210899P.
XX 20-FEB-2001; 2001US-0270216P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Jiang Y, Harlocker SL, Secret H;
XX WPI; 2002-114514/15.
XX
XX Novel isolated colon tumor polynucleotide differentially expressed in
XX colon tumor or colon metastatic tumor and polypeptides encoded by them,
XX useful for inhibiting development of cancer in patient.
XX
XX Claim 1; SEQ ID NO 963; 105pp; English.
XX
XX ABL36412 to ABL38645 represent human colon tumour antigen cDNA clones (I)
XX which were isolated from human colon tumour and colon metastatic tumour
XX cDNA libraries. (I) have cytostatic activity and can be used in vaccine
XX production. (I) can be used for stimulating and/or expanding T cells
XX specific for a tumour protein on contact with the T cells. They are also
XX useful for inhibiting the development of cancer in a patient. (I) can be
```

CC used as probes or primers for nucleic acid hybridisation, for preparing
 CC mutant species primers, or primers for use in genetic constructions. (I)
 CC can be used in the diagnosis of a colon tumour

SQ Sequence 542 BP; 163 A; 76 C; 111 G; 190 T; 0 U; 2 Other;

Query Match 82.2%; Score 14.8; DB 6; Length 542;

Best Local Similarity 88.9%; Pred. No. 2.6e+02;

Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGTGGGAGTTATCGAGC 18

DB 62 TGTGGAGAGTTATAGAC 79

RESULT 14

AAC52240/C

ID AAC52240 standard; DNA; 566 BP.

XX AAC52240;

AC AAC52240;

XX DT 18-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 70698.

XX Hybridisation assay; genetic mapping; gene expression control;

KW protein identification; signal transduction pathway; metabolic pathway;

KW promoter; termination sequence; ss.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-00301439.

XX 25-FEB-1999; 99US-0121825P.

PR 05-MAR-1999; 99US-0123180P.

PR 09-MAR-1999; 99US-0123548P.

PR 23-MAR-1999; 99US-0125788P.

PR 25-MAR-1999; 99US-0126284P.

PR 29-MAR-1999; 99US-0126785P.

PR 01-APR-1999; 99US-0127462P.

PR 06-APR-1999; 99US-0128234P.

PR 08-APR-1999; 99US-0128714P.

PR 16-APR-1999; 99US-0129845P.

PR 19-APR-1999; 99US-0130077P.

PR 21-APR-1999; 99US-0130449P.

PR 23-APR-1999; 99US-0130510P.

PR 23-APR-1999; 99US-0130891P.

PR 28-APR-1999; 99US-0131449P.

PR 30-APR-1999; 99US-0132048P.

PR 30-APR-1999; 99US-0132407P.

PR 04-MAY-1999; 99US-0132484P.

PR 05-MAY-1999; 99US-0132485P.

PR 06-MAY-1999; 99US-0132486P.

PR 06-MAY-1999; 99US-0132487P.

PR 07-MAY-1999; 99US-0132863P.

PR 11-MAY-1999; 99US-0134256P.

PR 14-MAY-1999; 99US-0134218P.

PR 14-MAY-1999; 99US-0134219P.

PR 14-MAY-1999; 99US-0134221P.

PR 14-MAY-1999; 99US-0134370P.

PR 18-MAY-1999; 99US-0134768P.

PR 19-MAY-1999; 99US-0134941P.

PR 20-MAY-1999; 99US-01351124P.

PR 21-MAY-1999; 99US-0135333P.

PR 24-MAY-1999; 99US-0135629P.

PR 25-MAY-1999; 99US-0136021P.

PR 27-MAY-1999; 99US-0136382P.

PR 28-MAY-1999; 99US-0136782P.

PR 01-JUN-1999; 99US-0137222P.

PR 03-JUN-1999; 99US-0137528P.

PR 04-JUN-1999; 99US-0137502P.

PR 07-JUN-1999; 99US-0137724P.

PR 08-JUN-1999; 99US-0138094P.

PR 10-JUN-1999; 99US-0138540P.

PR 10-JUN-1999; 99US-0138847P.

PR 14-JUN-1999; 99US-0139119P.

PR 16-JUN-1999; 99US-0139452P.

PR 17-JUN-1999; 99US-0139453P.

PR 17-JUN-1999; 99US-0139492P.

PR 18-JUN-1999; 99US-0139454P.

PR 18-JUN-1999; 99US-0139455P.

PR 18-JUN-1999; 99US-0139456P.

PR 18-JUN-1999; 99US-0139457P.

PR 18-JUN-1999; 99US-0139458P.

PR 18-JUN-1999; 99US-0139459P.

PR 18-JUN-1999; 99US-0139460P.

PR 18-JUN-1999; 99US-0139461P.

PR 18-JUN-1999; 99US-0139462P.

PR 18-JUN-1999; 99US-0139463P.

PR 18-JUN-1999; 99US-0139750P.

PR 18-JUN-1999; 99US-0139763P.

PR 21-JUN-1999; 99US-0139817P.

PR 22-JUN-1999; 99US-0139899P.

PR 23-JUN-1999; 99US-0140353P.

PR 23-JUN-1999; 99US-0140354P.

PR 24-JUN-1999; 99US-0140695P.

PR 28-JUN-1999; 99US-0140823P.

PR 29-JUN-1999; 99US-0140991P.

PR 30-JUN-1999; 99US-0141287P.

PR 01-JUL-1999; 99US-0141842P.

PR 01-JUL-1999; 99US-0142154P.

PR 02-JUL-1999; 99US-0142055P.

PR 06-JUL-1999; 99US-0142390P.

PR 08-JUL-1999; 99US-0142803P.

PR 09-JUL-1999; 99US-0142920P.

PR 12-JUL-1999; 99US-0142977P.

PR 13-JUL-1999; 99US-0143542P.

PR 14-JUL-1999; 99US-0143624P.

PR 15-JUL-1999; 99US-0144005P.

PR 16-JUL-1999; 99US-0144085P.

PR 16-JUL-1999; 99US-0144086P.

PR 19-JUL-1999; 99US-0144325P.

PR 19-JUL-1999; 99US-0144331P.

PR 19-JUL-1999; 99US-0144332P.

PR 19-JUL-1999; 99US-0144333P.

PR 19-JUL-1999; 99US-0144334P.

PR 19-JUL-1999; 99US-0144335P.

PR 20-JUL-1999; 99US-0144352P.

PR 20-JUL-1999; 99US-0144632P.

PR 20-JUL-1999; 99US-0144884P.

PR 21-JUL-1999; 99US-0144814P.

PR 21-JUL-1999; 99US-0145086P.

PR 21-JUL-1999; 99US-0145088P.

PR 22-JUL-1999; 99US-0145085P.

PR 22-JUL-1999; 99US-0145087P.

PR 22-JUL-1999; 99US-0145089P.

PR 22-JUL-1999; 99US-0145192P.

PR 23-JUL-1999; 99US-0145145P.

PR 23-JUL-1999; 99US-0145218P.

PR 23-JUL-1999; 99US-0145224P.

PR 26-JUL-1999; 99US-0145276P.

PR 27-JUL-1999; 99US-0145913P.

PR 27-JUL-1999; 99US-0145918P.

PR 27-JUL-1999; 99US-0145919P.

PR 28-JUL-1999; 99US-0145951P.

PR 02-AUG-1999; 99US-0146386P.

PR 02-AUG-1999; 99US-0146388P.

PR 03-AUG-1999; 99US-0146389P.

PR 04-AUG-1999; 99US-0147038P.

PR 04-AUG-1999; 99US-0147204P.

PR 05-AUG-1999; 99US-0147302P.

PR 05-AUG-1999; 99US-0147192P.

```
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147316P.
PR 09-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0149368P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 25-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 28-SEP-1999; 99US-0156596P.
PR 29-SEP-1999; 99US-0157117P.
PR 04-OCT-1999; 99US-0157753P.
PR 05-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159633P.
PR 14-OCT-1999; 99US-0159637P.
PR 18-OCT-1999; 99US-0159638P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 82.2%; Score 14.8; DB 3; Length 566;

Best Local Similarity 88.9%; Pred. No. 2.6e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 TGTGGGAGTTATCGAC 18
DB 61 TTGTGTGAGTTATCGAC 44

RESULT 15
AAC54690/c
ID AAC54690 standard; DNA, 566 BP.
XX
AC AAC54690;
XX
XX 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 78737.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
XX 05-MAR-1999; 99US-0123180P.
XX 09-MAR-1999; 99US-0123548P.
XX 23-MAR-1999; 99US-0125788P.
XX 25-MAR-1999; 99US-0126264P.
XX 01-APR-1999; 99US-0127462P.
XX 06-APR-1999; 99US-0128234P.
XX 08-APR-1999; 99US-0128714P.
XX 16-APR-1999; 99US-0129845P.
XX 19-APR-1999; 99US-0130077P.
XX 21-APR-1999; 99US-0130449P.
XX 23-APR-1999; 99US-0130510P.
XX 28-APR-1999; 99US-0130891P.
XX 30-APR-1999; 99US-0132048P.
XX 04-MAY-1999; 99US-0132407P.
XX 05-MAY-1999; 99US-0132484P.
XX 05-MAY-1999; 99US-0132485P.
XX 06-MAY-1999; 99US-0132486P.
XX 07-MAY-1999; 99US-0132487P.
XX 11-MAY-1999; 99US-0132863P.
XX 14-MAY-1999; 99US-0134218P.
XX 14-MAY-1999; 99US-0134219P.
XX 14-MAY-1999; 99US-0134221P.
XX 18-MAY-1999; 99US-0134370P.
XX 19-MAY-1999; 99US-0134768P.
XX 20-MAY-1999; 99US-0134941P.
XX 21-MAY-1999; 99US-0135124P.
XX 24-MAY-1999; 99US-0135533P.
XX 24-MAY-1999; 99US-0135629P.
XX 25-MAY-1999; 99US-0136021P.
XX 27-MAY-1999; 99US-0136392P.
XX 28-MAY-1999; 99US-0136782P.
XX 01-JUN-1999; 99US-0137222P.
XX 03-JUN-1999; 99US-0137528P.
XX 04-JUN-1999; 99US-0137502P.
XX 07-JUN-1999; 99US-0137724P.
XX 08-JUN-1999; 99US-0138094P.
XX 10-JUN-1999; 99US-0138540P.
XX 10-JUN-1999; 99US-0138847P.
XX 14-JUN-1999; 99US-0139119P.
```

```
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139751P.
PR 21-JUN-1999; 99US-0139817P.
PR 21-JUN-1999; 99US-0139899P.
PR 22-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 23-JUN-1999; 99US-0140355P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140696P.
PR 28-JUN-1999; 99US-0140697P.
PR 28-JUN-1999; 99US-0140698P.
PR 28-JUN-1999; 99US-0140699P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142155P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144332P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144804P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.

PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149388P.
PR 17-AUG-1999; 99US-0149417P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 30-AUG-1999; 99US-0151080P.
PR 31-AUG-1999; 99US-0151303P.
PR 01-SEP-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-015753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159233P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 18-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 82.2%; Score 14.8; DB 3; Length 566;
Best Local Similarity 88.9%; Pred. NO. 2.6e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 TCTGGGGAGTTATCGAC 18
DB 61 TTGTGTGAGTTATCGAC 44
```

Search completed: April 24, 2004, 18:45:05
Job time : 155.222 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2004, 18:26:15 ; Search time 33.3333 Seconds
(without alignments)
299.673 Million cell updates/sec

Title: US-10-084-555A-117

Perfect score: 18
Sequence: 1 tgtggggagttatcgagc 18

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/ptodata/2/ina/5A COMB.seq:*
- 2: /cgn2_6/ptodata/2/ina/5B COMB.seq:*
- 3: /cgn2_6/ptodata/2/ina/6A COMB.seq:*
- 4: /cgn2_6/ptodata/2/ina/6B COMB.seq:*
- 5: /cgn2_6/ptodata/2/ina/PCTUS COMB.seq:*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
C 1	14.8	82.2	1830121	4	US-09-557-884-1
C 2	14.8	82.2	1830121	4	US-09-543-990A-1
C 3	14.4	80.0	6415	4	US-08-961-527-162
C 4	14	77.8	1664976	4	US-08-916-421B-1
5	13.8	76.7	357	4	US-09-489-039A-5765
6	13.8	76.7	873	4	US-08-484-841A-7
7	13.8	76.7	1158	3	US-08-591-605-1
8	13.8	76.7	1159	3	US-08-422-669B-3
9	13.8	76.7	1159	4	US-09-124-934A-3
10	13.8	76.7	1159	4	US-08-334-251D-3
11	13.8	76.7	1692	4	US-09-600-991-3
12	13.8	76.7	1709	4	US-09-600-991-21
13	13.8	76.7	1725	4	US-09-600-991-1
14	13.8	76.7	1759	4	US-09-600-991-22
15	13.8	76.7	2172	3	US-08-030-410-2
16	13.8	76.7	2172	4	US-09-600-991-17
17	13.8	76.7	2184	1	US-07-815-333A-1
18	13.8	76.7	2288	1	US-08-290-937B-4
19	13.8	76.7	2289	1	US-07-838-410-2
20	13.8	76.7	2635	4	US-09-484-617-3
21	13.8	76.7	2635	4	US-09-023-555-1446
22	13.8	76.7	5175	2	US-08-843-530B-3
23	13.8	76.7	5697	2	US-08-843-530B-5
24	13.8	76.7	6545	2	US-08-843-530B-1
25	13.8	76.7	12665	4	US-08-961-527-134
26	13.8	76.7	36741	3	US-09-301-665-3
C 27	13.8	76.7	42157	4	US-08-311-731A-126

28 13.8 76.7 4403765 3 US-09-103-840A-2 Sequence 2, Appli
29 13.8 76.7 4411529 3 US-09-103-840A-1 Sequence 1, Appli
30 13.4 74.4 278 4 US-09-702-705-1032 Sequence 1032, Ap
31 13.4 74.4 278 4 US-09-736-457-1032 Sequence 1032, Ap
32 13.4 74.4 278 4 US-09-614-124B-1032 Sequence 1032, Ap
33 13.4 74.4 278 4 US-09-671-325-1032 Sequence 1032, Ap
34 13.4 74.4 681 4 US-09-149-476-189 Sequence 189, App
35 13.4 74.4 787 4 US-09-401-064-132 Sequence 132, App
36 13.4 74.4 930 4 US-09-227-357-61 Sequence 61, Appl
37 13.4 74.4 930 4 US-09-227-357-61 Sequence 146, App
38 13.4 74.4 1406 2 US-08-933-750C-72 Sequence 72, Appl
39 13.4 74.4 1406 3 US-09-234-613-72 Sequence 1180, Ap
40 13.4 74.4 1545 4 US-09-543-881A-1180 Sequence 521, App
41 13.4 74.4 1665 4 US-09-489-039A-521 Sequence 62, Appl
42 13.4 74.4 3213 4 US-09-107-532A-62 Sequence 10, Appl
43 13.4 74.4 3466 3 US-09-027-166-10 Sequence 166, App
44 13.4 74.4 5406 4 US-08-961-527-166 Sequence 27, Appl
45 13.4 74.4 9100 2 US-08-743-637B-27

ALIGNMENTS

RESULT 1
US-09-557-884-1/c
; Sequence 1, Application US/09557884
; Patent No. 6506591
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/557,884
; FILING DATE: 25-Apr-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,102
; FILING DATE: JUN-5-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB186P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Query Match 82.2%; Score 14.8; DB 4; Length 1830121;
Best Local Similarity 88.9%; Pred. No. 92;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 TGTGGGGAGTTATCGAGC 18

Db 1438046 TCTGGGGAGTTGTCGAGC 1438029

RESULT 2

US-09-643-990A-1/c
; Sequence 1, Application US/09643990A
; Patent No. 6528289
; GENERAL INFORMATION:
; APPLICANT: Robert D. Fleischmann
; Mark D. Adams
; Owen White
; Hamilton O. Smith
; J. Craig Venter
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville,
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/643,990A
; FILING DATE: 23-Aug-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,429
; FILING DATE: 1995-06-07
; APPLICATION NUMBER: 08/426,787
; FILING DATE: 1995-04-21
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB186P1C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-610-5790
; TELEFAX: 310-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-643-990A-1
Query Match 82.2%; Score 14.8; DB 4; Length 1830121;
Best Local Similarity 88.9%; Pred. No. 92;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 TGTGGGGAGTTATCGAGC 18
Db 1438046 TCTGGGGAGTTGTCGAGC 1438029
RESULT 3
US-08-961-527-162/c
; Sequence 162, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kinsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 162:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6415 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-961-527-162

Query Match 80.0%; Score 14.4; DB 4; Length 6415;
Best Local Similarity 93.8%; Pred. No. 74;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 TGGGGAGTTATCGAGC 18

Db 3810 TGAGGAGTTATCGAGC 3795

RESULT 4

US-08-916-421B-1/c
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanoco
; Patent No. 6503729
; TITLE OF INVENTION: jannaschii
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature

LOCATION: (84808)..(84808)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84812)..(84812)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98120)..(98120)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98159)..(98159)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98239)..(98239)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98266)..(98266)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98343)..(98343)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (103998)..(103998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (148948)..(148948)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (163385)..(163385)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (191989)..(191989)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (191995)..(191995)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (231980)..(231980)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234187)..(234187)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234220)..(234220)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234814)..(234814)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309398)..(309398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309418)..(309418)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312837)..(312837)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312993)..(312993)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (319226)..(319226)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (559167)..(559167)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (559241)..(559241)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (600592)..(600592)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (622708)..(622708)

OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (657081)..(657081)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (657203)..(657203)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (674435)..(674435)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (682442)..(682442)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (713652)..(713652)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (741684)..(741684)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (779455)..(779455)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (779676)..(779676)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (858539)..(858539)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (871619)..(871619)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1084830)..(1084830)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1096846)..(1096846)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1119881)..(1119881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1130881)..(1130881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1310988)..(1310988)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1313224)..(1313224)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1349473)..(1349473)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1349491)..(1349491)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1470091)..(1470091)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1569020)..(1569020)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1602912)..(1602912)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1603734)..(1603734)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1637998)..(1637998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1664855)..(1664855)
OTHER INFORMATION: n equals a, t, c, or g

US-08-916-421B-1

Query Match 77.8%; Score 14; DB 4; Length 1664976;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGGGGAGTTATC 14
|||
DB 402613 TGTGGGGAGTTATC 402600

RESULT 5

US-09-489-039A-5765
; Sequence 765, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 5765
; LENGTH: 357
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-5765

Query Match 76.7%; Score 13.8; DB 4; Length 357;
Best Local Similarity 88.2%; Pred. No. 1e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGTGGGGAGTTATCGAG 17
|||
DB 149 TGTGGGGAGTTAGCGAG 165

RESULT 6

US-08-484-841A-7
; Sequence 7, Application US/08484841A
; Patent No. 6566098
; GENERAL INFORMATION:

; APPLICANT: Chan, Andrew ML
; APPLICANT: Rubin, Jeffrey L.
; APPLICANT: Bottaro, Donald P.
; APPLICANT: Aaronson, Stuart A.
; APPLICANT: Stahl, Stephen J.
; APPLICANT: Wingfield, Paul T.
; APPLICANT: Cioce, Vittoria
; TITLE OF INVENTION: TRUNCATED HEPATOCYTE GROWTH FACTOR
; TITLE OF INVENTION: VARIANTS
; FILE REFERENCE: 14014.0269
; CURRENT APPLICATION NUMBER: US/08/484,841A
; CURRENT FILING DATE: 2002-03-13
; PRIOR APPLICATION NUMBER: 08/130,134
; PRIOR FILING DATE: 1993-10-04
; PRIOR APPLICATION NUMBER: 07/655,502
; PRIOR FILING DATE: 1991-02-15
; PRIOR APPLICATION NUMBER: 07/582,063
; PRIOR FILING DATE: 1990-09-14
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 873
; TYPE: DNA

; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Artificial Sequence; / No. 6566098e = synthetic construct
; NAME/KEY: CDS
; LOCATION: (1)...(873)

US-08-484-841A-7

Query Match 76.7%; Score 13.8; DB 4; Length 873;
Best Local Similarity 88.2%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGTGGGGAGTTATCGAG 17
|||
DB 645 TGTGGGGAGTTATCGAG 661

RESULT 7

US-08-591-605-1
; Sequence 1, Application US/08591605
; Patent No. 6060238
; GENERAL INFORMATION:
; APPLICANT: Dixit, Vishva M.
; TITLE OF INVENTION: METHOD AND COMPOSITION FOR REGULATING
; TITLE OF INVENTION: APOPTOSIS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: PALO ALTO
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/591,605
; FILING DATE: 09-FEB-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: KOSKI, ANTOINETTE F.
; REGISTRATION NUMBER: 34,202
; REFERENCE/DOCKET NUMBER: 20344-21036.21
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 813-5600
; TELEFAX: (415) 494-0792
; TELEX: 706141 MRSNFOERS SFO
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1158 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 224..1054
; OTHER INFORMATION: /product= "Yama peptide sequence"

; FEATURE:
; NAME/KEY: sig peptide
; LOCATION: 704..718
; OTHER INFORMATION: /note= "region coding for
; OTHER INFORMATION: pentapeptide, wherein cysteine is catalytic"
; FEATURE:
; NAME/KEY: mutation
; LOCATION: replace(710..712, "atg")
; OTHER INFORMATION: /note= "Mutation of catalytic cysteine to methionine"
; FEATURE:
; NAME/KEY: mutation
; LOCATION: replace(710..712, "gcc")
; OTHER INFORMATION: /note= "mutation of catalytic cysteine to alanine"
; FEATURE:
; NAME/KEY: mutation
; LOCATION: replace(710..712, "ggc")
; OTHER INFORMATION: /note= "mutation of catalytic cysteine to alanine"
; FEATURE:
; NAME/KEY: mutation
; LOCATION: replace(710..712, "ggc")
; OTHER INFORMATION: /note= "mutation of catalytic cysteine to alanine"
; FEATURE:
; NAME/KEY: mutation

```
; LOCATION: replace(710..712, "gct")
; OTHER INFORMATION: /note= "mutation of catalytic cysteine to alanine"
; FEATURE:
; NAME/KEY: mutation
; LOCATION: replace(710..712, "gca")
; OTHER INFORMATION: /note= "mutation of catalytic cysteine to alanine"
; US-08-591-605-1
Query Match      76.7%; Score 13.8; DB 3; Length 1158;
Best Local Similarity 88.2%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 GTGGGGAGTTATCGAGC 18
      ||| ||||| ||||| |||||
DB      125 GTGAGGAGTTAGCGAGC 141

RESULT 8
US-08-462-969B-3
; Sequence 3, Application US/08462969B
; Patent No. 6087150
; GENERAL INFORMATION:
; APPLICANT: He, Wei-Wu et al.
; TITLE OF INVENTION: Interleukin-1 Beta Converting Enzyme
; FILE REFERENCE: PF140P1
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Ave.
; CITY: Rockville
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,969B
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/334,251
; FILING DATE: 11-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PF140P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8439
; INFORMATION FOR SEQ ID NO: 3:
; LENGTH: 1159 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-462-969B-3
Query Match      76.7%; Score 13.8; DB 3; Length 1159;
Best Local Similarity 88.2%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 GTGGGGAGTTATCGAGC 18
      ||| ||||| ||||| |||||
DB      124 GTGAGGAGTTAGCGAGC 140

RESULT 9
US-09-124-934A-3
; Sequence 3, Application US/09124934A
```

```
; Patent No. 6495519
; GENERAL INFORMATION:
; APPLICANT: He, Wei-Wu et al.
; TITLE OF INVENTION: Interleukin-1 Beta Converting Enzyme Like Apoptosis Protease 3 a
; FILE REFERENCE: PF140C1
; CURRENT APPLICATION NUMBER: US/09/124,934A
; CURRENT FILING DATE: 1994-11-01
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 1159
; TYPE: DNA
; ORGANISM: homo sapiens
; US-09-124-934A-3
Query Match      76.7%; Score 13.8; DB 4; Length 1159;
Best Local Similarity 88.2%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 GTGGGGAGTTATCGAGC 18
      ||| ||||| ||||| |||||
DB      124 GTGAGGAGTTAGCGAGC 140

RESULT 10
US-08-334-251D-3
; Sequence 3, Application US/08334251D
; Patent No. 6538121
; GENERAL INFORMATION:
; APPLICANT: He et al.
; TITLE OF INVENTION: Interleukin-1 Beta Converting Enzyme Like Apoptosis Protease 3 a
; FILE REFERENCE: PF140
; CURRENT APPLICATION NUMBER: US/08/334,251D
; CURRENT FILING DATE: 1994-11-01
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 1159
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-08-334-251D-3
Query Match      76.7%; Score 13.8; DB 4; Length 1159;
Best Local Similarity 88.2%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 GTGGGGAGTTATCGAGC 18
      ||| ||||| ||||| |||||
DB      124 GTGAGGAGTTAGCGAGC 140

RESULT 11
US-09-600-991-3
; Sequence 3, Application US/09600991
; Patent No. 6551991
; GENERAL INFORMATION:
; APPLICANT: MEDICO, Enzo
; APPLICANT: MICHELLE, Paolo
; APPLICANT: COLLESI, Chiara
; APPLICANT: CASELLI, Gianfranco
; APPLICANT: COMOGLIO, Paolo
; TITLE OF INVENTION: RECOMBINANT PROTEINS DERIVED FROM HGF AND MSP
; FILE REFERENCE: 0471-0162P
; CURRENT APPLICATION NUMBER: US/09/600,991
; CURRENT FILING DATE: 2001-08-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1692
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Metron F-1 DNA coding sequence
```

US-09-600-991-3

Query Match 76.7%; Score 13.8; DB 4; Length 1692;
Best Local Similarity 88.2%; Pred. No. 1.3e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGTGGGAGATTATCGAG 17

DB 1422 TGGGAGAGTTATCGAG 1438

RESULT 12

US-09-600-991-21

; Sequence 21, Application US/09600991

; Patent No. 6551991

; GENERAL INFORMATION:

; APPLICANT: MEDICO, Enzo

; APPLICANT: MICHIELI, Paolo

; APPLICANT: COLLESI, Chiara

; APPLICANT: CASELLI, Gianfranco

; APPLICANT: COMOGGIO, Paolo

; TITLE OF INVENTION: RECOMBINANT PROTEINS DERIVED FROM HGF AND MSP

; FILE REFERENCE: 0471-0162P

; CURRENT APPLICATION NUMBER: US/09/600,991

; CURRENT FILING DATE: 2001-08-20

; NUMBER OF SEQ ID NOS: 22

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 21

; LENGTH: 1709

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Metron F-1

US-09-600-991-21

Query Match

Best Local Similarity 76.7%; Score 13.8; DB 4; Length 1709;

; Sequence 1, Application US/09600991

; Patent No. 6551991

; GENERAL INFORMATION:

; APPLICANT: MEDICO, Enzo

; APPLICANT: MICHIELI, Paolo

; APPLICANT: COLLESI, Chiara

; APPLICANT: CASELLI, Gianfranco

; APPLICANT: COMOGGIO, Paolo

; TITLE OF INVENTION: RECOMBINANT PROTEINS DERIVED FROM HGF AND MSP

; FILE REFERENCE: 0471-0162P

; CURRENT APPLICATION NUMBER: US/09/600,991

; CURRENT FILING DATE: 2001-08-20

; NUMBER OF SEQ ID NOS: 22

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 1

; LENGTH: 1725

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Magic F-1 DNA coding sequence

US-09-600-991-1

Query Match 76.7%; Score 13.8; DB 4; Length 1725;

Best Local Similarity 88.2%; Pred. No. 1.3e+02;

Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGTGGGAGATTATCGAG 17

DB 630 TGGGAGAGTTATCGAG 646

RESULT 14

US-09-600-991-22

; Sequence 22, Application US/09600991

; Patent No. 6551991

; GENERAL INFORMATION:

; APPLICANT: MEDICO, Enzo

; APPLICANT: MICHIELI, Paolo

; APPLICANT: COLLESI, Chiara

; APPLICANT: CASELLI, Gianfranco

; APPLICANT: COMOGGIO, Paolo

; TITLE OF INVENTION: RECOMBINANT PROTEINS DERIVED FROM HGF AND MSP

; FILE REFERENCE: 0471-0162P

; CURRENT APPLICATION NUMBER: US/09/600,991

; CURRENT FILING DATE: 2001-08-20

; NUMBER OF SEQ ID NOS: 22

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 22

; LENGTH: 1759

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Magic F-1

US-09-600-991-22

Query Match 76.7%; Score 13.8; DB 4; Length 1759;

Best Local Similarity 88.2%; Pred. No. 1.3e+02;

Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGTGGGAGATTATCGAG 17

DB 657 TGGGAGAGTTATCGAG 673

RESULT 15

US-08-030-410-2

; Sequence 2, Application US/08030410

; Patent No. 6221359

; GENERAL INFORMATION:

; APPLICANT: Koniya, Atsushi

; APPLICANT: Nakahata, Tatsutoshi

; APPLICANT: Kubo, Tetsuo

; APPLICANT: Tanaka, Ryuei

; APPLICANT: Kawano, Genji

; APPLICANT: Sudo, Tetsuo

; APPLICANT: Sano, Emiko

; APPLICANT: Kojima, Katsuki

; TITLE OF INVENTION: HEMOPOIETIC STEM CELL MULTIPLIER

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

ADDRESSEE: Scully, Scott, Murphy & Presser

STREET: 400 Garden City Plaza

CITY: Garden City

STATE: New York

COUNTRY: United States

ZIP: 11530

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/030,410

FILING DATE: 19930521

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Digiglio, Frank S.

REGISTRATION NUMBER: 31,346

REFERENCE/DOCKET NUMBER: 8898

TELECOMMUNICATION INFORMATION:

```

; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2172 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2169
; US-08-030-410-2

```

```

Query Match      76.7%; Score 13.8; DB 3; Length 2172;
Best Local Similarity 88.2%; Pred. No. 1.4e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy      1 TGTGGGAGTTATCGAG 17
Db      630 TGGGGAGAGTTATCGAG 646

```

Search completed: April 24, 2004, 21:13:48
Job time : 45.3333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2004, 19:25:41 ; Search time 139.111 Seconds
(without alignments)
583.385 Million cell updates/sec

Title: US-10-084-555A-117

Perfect score: 18

Sequence: 1 tgtggggaggtatcgagc 18

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2907579 seqs, 2254313464 residues

Total number of hits satisfying chosen parameters: 5815158

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:

```

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	16.4	91.1	65559	13	US-10-087-192-841
2	15.4	85.6	6063	15	US-10-240-453-268
3	15.4	85.6	19653	15	US-10-311-455-1308
4	15.4	85.6	263744	15	US-10-229-834A-6
5	15	83.3	3337	10	US-09-586-480-50
6	15	83.3	3337	10	US-09-895-298-51
7	14.8	82.2	468	10	US-09-918-995-24221
8	14.8	82.2	539	9	US-09-878-178-963
9	14.8	82.2	539	14	US-10-046-935-963
10	14.8	82.2	539	15	US-10-146-502-963
11	14.8	82.2	738	9	US-09-738-626-2293
12	14.8	82.2	1058	13	US-10-027-632-257422
13	14.8	82.2	1058	13	US-10-027-632-257423
14	14.8	82.2	1058	16	US-10-027-632-257422

15	14.8	82.2	1058	16	US-10-027-632-257423	Sequence 257423,
16	14.8	82.2	1214	13	US-10-424-599-66564	Sequence 66564, A
17	14.8	82.2	1527	15	US-10-369-493-26395	Sequence 26395, A
18	14.8	82.2	1635	9	US-09-728-952-22	Sequence 22, Appl
19	14.8	82.2	2193	13	US-10-282-122A-22315	Sequence 22315, A
20	14.8	82.2	3361	10	US-09-977-418-19	Sequence 19, Appl
21	14.8	82.2	3361	10	US-09-977-033A-19	Sequence 19, Appl
22	14.8	82.2	3361	10	US-09-977-751C-19	Sequence 19, Appl
23	14.8	82.2	3361	10	US-09-977-639A-19	Sequence 19, Appl
24	14.8	82.2	3361	11	US-09-977-819B-19	Sequence 19, Appl
25	14.8	82.2	3671	13	US-10-147-493-141	Sequence 141, App
26	14.8	82.2	3671	13	US-10-145-127-141	Sequence 141, App
27	14.8	82.2	3671	13	US-10-160-503-141	Sequence 141, App
28	14.8	82.2	3671	13	US-10-143-118-141	Sequence 141, App
29	14.8	82.2	3671	13	US-10-144-993-141	Sequence 141, App
30	14.8	82.2	3671	13	US-10-158-787-141	Sequence 141, App
31	14.8	82.2	3671	13	US-10-140-024-141	Sequence 141, App
32	14.8	82.2	3671	13	US-10-140-808-141	Sequence 141, App
33	14.8	82.2	3671	13	US-10-152-405-141	Sequence 141, App
34	14.8	82.2	3671	13	US-10-127-852A-141	Sequence 141, App
35	14.8	82.2	3671	13	US-10-127-900A-141	Sequence 141, App
36	14.8	82.2	3671	13	US-10-128-685A-141	Sequence 141, App
37	14.8	82.2	3671	13	US-10-131-820A-141	Sequence 141, App
38	14.8	82.2	3671	13	US-10-142-886-141	Sequence 141, App
39	14.8	82.2	3671	13	US-10-146-728-141	Sequence 141, App
40	14.8	82.2	3671	13	US-10-146-786-141	Sequence 141, App
41	14.8	82.2	3671	13	US-10-147-499-141	Sequence 141, App
42	14.8	82.2	3671	13	US-10-157-798-141	Sequence 141, App
43	14.8	82.2	3671	15	US-10-028-072-141	Sequence 141, App
44	14.8	82.2	3671	15	US-10-121-049-141	Sequence 141, App
45	14.8	82.2	3671	15	US-10-123-904-141	Sequence 141, App

ALIGNMENTS

RESULT 1

US-10-087-192-841
; Sequence 841, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 841
; LENGTH: 65559
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(65559)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-841

Query Match 91.1%; Score 16.4; DB 13; Length 65559;
Best Local Similarity 94.4%; Pred. No. 44;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGTGGGAGTTATCGAC 18

DB 25463 TGTGGGAGTTATCAAGC 25480

```
RESULT 2
US-10-240-453-268
; Sequence 268, Application US/10240453
; Publication No. US20030148326A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
; TITLE OF INVENTION: Transcription
; TITLE OF INVENTION: by Means of Assessing the Methylation Status of Genes Associated
; TITLE OF INVENTION: With DNA Transcription
; FILE REFERENCE: 5013.1009
; CURRENT APPLICATION NUMBER: US/10/240.453
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03973
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 350
; SEQ ID NO 268
; LENGTH: 6063
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-453-268

Query Match      85.6%; Score 15.4; DB 15; Length 6063;
Best Local Similarity 94.1%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGTGGGGAGTTATCGAG 17
DB 489 TGAGGGGAGTTATCGAG 505

RESULT 3
US-10-311-455-1308
; Sequence 1308, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; TITLE OF INVENTION: Cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311.455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1308
; LENGTH: 19653
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1308

Query Match      85.6%; Score 15.4; DB 15; Length 19653;
Best Local Similarity 94.1%; Pred. No. 1.5e+02;
```

```
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGTGGGGAGTTATCGAG 17
DB 17158 TGTGGGGAGTTATAGAG 17174

RESULT 4
US-10-229-834A-6
; Sequence 6, Application US/10229834A
; Publication No. US20030150003A1
; GENERAL INFORMATION:
; APPLICANT: Lawrence Berkeley National Laboratory
; APPLICANT: Rubin, Edward
; APPLICANT: Pennacchio, Len
; TITLE OF INVENTION: A novel apolipoprotein gene involved in lipid metabolism
; FILE REFERENCE: IB-1709
; CURRENT APPLICATION NUMBER: US/10/229.834A
; CURRENT FILING DATE: 2002-08-27
; PRIOR APPLICATION NUMBER: US 60/318,219
; PRIOR FILING DATE: 2001-08-27
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 263744
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (750)..(850)
; OTHER INFORMATION: gap of unknown length
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1584)..(1683)
; OTHER INFORMATION: gap of unknown length
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2154)..(2154)
; OTHER INFORMATION: gap of unknown length
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2358)..(2457)
; OTHER INFORMATION: gap of unknown length
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3208)..(3307)
; OTHER INFORMATION: gap of unknown length
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4044)..(4143)
; OTHER INFORMATION: gap of unknown length
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4809)..(4908)
; OTHER INFORMATION: gap of unknown length
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4921)..(4921)
; OTHER INFORMATION: gap of unknown length
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5652)..(5751)
; OTHER INFORMATION: gap of unknown length
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6469)..(6469)
; OTHER INFORMATION: gap of unknown length
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6477)..(6477)
; OTHER INFORMATION: gap of unknown length
; FEATURE:
; NAME/KEY: misc_feature
```



```
, LOCATION: (6482)..(6482)
, OTHER INFORMATION: gap of unknown length
, FEATURE:
, NAME/KEY: misc feature
, LOCATION: (6487)..(6487)
, OTHER INFORMATION: gap of unknown length
, FEATURE:
, NAME/KEY: misc feature
, LOCATION: (6489)..(6490)
, OTHER INFORMATION: gap of unknown length
, FEATURE:
, NAME/KEY: misc feature
, LOCATION: (6523)..(6522)
, OTHER INFORMATION: gap of unknown length
, FEATURE:
, NAME/KEY: misc feature
, LOCATION: (6637)..(6637)
, OTHER INFORMATION: gap of unknown length
, FEATURE:
, NAME/KEY: misc feature
, LOCATION: (7410)..(7509)
, OTHER INFORMATION: gap of unknown length
, FEATURE:
, NAME/KEY: misc feature
, LOCATION: (8177)..(8177)
, OTHER INFORMATION: gap of unknown length
, FEATURE:
, NAME/KEY: misc feature
, LOCATION: (8248)..(8347)
, OTHER INFORMATION: gap of unknown length
, FEATURE:
, NAME/KEY: misc feature
, LOCATION: (8358)..(8358)
, OTHER INFORMATION: gap of unknown length
, FEATURE:
, NAME/KEY: misc feature
, LOCATION: (8967)..(8967)
, OTHER INFORMATION: gap of unknown length
, FEATURE:
, NAME/KEY: misc feature
, LOCATION: (9057)..(9057)
, OTHER INFORMATION: gap of unknown length
, FEATURE:
, NAME/KEY: misc feature
, LOCATION: (9105)..(9204)
, OTHER INFORMATION: gap of unknown length
, FEATURE:
, NAME/KEY: misc feature
, LOCATION: (9980)..(10079)
, OTHER INFORMATION: gap of unknown length
, FEATURE:
, NAME/KEY: misc feature
, LOCATION: (10032)..(10092)
, OTHER INFORMATION: gap of unknown length
, FEATURE:
, NAME/KEY: misc feature
, LOCATION: (10827)..(10926)
, OTHER INFORMATION: gap of unknown length
, FEATURE:
, NAME/KEY: misc feature
, LOCATION: (11701)..(11800)
, OTHER INFORMATION: gap of unknown length
, FEATURE:
, NAME/KEY: misc feature
, LOCATION: (11816)..(11816)
, OTHER INFORMATION: gap of unknown length
, FEATURE:
, NAME/KEY: misc feature
, LOCATION: (11827)..(11827)
, OTHER INFORMATION: gap of unknown length
, FEATURE:
, NAME/KEY: misc feature
, LOCATION: (11829)..(11829)
```

```
, OTHER INFORMATION: gap of unknown length
, FEATURE:
, NAME/KEY: misc feature
, LOCATION: (12555)..(12664)
, OTHER INFORMATION: gap of unknown length
, FEATURE:
, NAME/KEY: misc feature
, LOCATION: (13327)..(13327)
, OTHER INFORMATION: gap of unknown length
, FEATURE:
, NAME/KEY: misc feature
, LOCATION: (13440)..(13539)
, OTHER INFORMATION: gap of unknown length
, FEATURE:
, NAME/KEY: misc feature
, LOCATION: (13554)..(13554)
, OTHER INFORMATION: gap of unknown length
, FEATURE:
, NAME/KEY: misc feature
, LOCATION: (14133)..(14133)
, OTHER INFORMATION: gap of unknown length
, FEATURE:
, NAME/KEY: misc feature
, LOCATION: (14213)..(14213)
, OTHER INFORMATION: gap of unknown length
, FEATURE:
, NAME/KEY: misc feature
, LOCATION: (14293)..(14392)
, OTHER INFORMATION: gap of unknown length
, FEATURE:
, NAME/KEY: misc feature
, LOCATION: (14559)..(14559)
, OTHER INFORMATION: gap of unknown length
, FEATURE:
, NAME/KEY: misc feature
, LOCATION: (14900)..(14900)
, OTHER INFORMATION: gap of unknown length
, FEATURE:
, NAME/KEY: misc feature
, LOCATION: (15119)..(15218)
, OTHER INFORMATION: gap of unknown length
, FEATURE:
, NAME/KEY: misc feature
, LOCATION: (15231)..(15233)
, OTHER INFORMATION: gap of unknown length
, FEATURE:
, NAME/KEY: misc feature
, LOCATION: (15919)..(15919)
, OTHER INFORMATION: gap of unknown length
, FEATURE:
, NAME/KEY: misc feature
, LOCATION: (15922)..(15922)
, OTHER INFORMATION: gap of unknown length
, FEATURE:
, NAME/KEY: misc feature
, LOCATION: (15962)..(16061)
, OTHER INFORMATION: gap of unknown length
, FEATURE:
, NAME/KEY: misc feature
, LOCATION: (16757)..(16757)
, OTHER INFORMATION: gap of unknown length
, FEATURE:
, NAME/KEY: misc feature
```

Query Match 85.6%; Score 15.4; DB 15; Length 263744;
Best Local Similarity 94.1%; Pred. No. 1.7e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GTGGGAGCTTATCGAGC 18
DB 157885 GTGGGAGGTATCGAGC 157901

```
RESULT 5
US-09-986-480-50/c
; Sequence 50, Application US/09986480
; Publication No. US20030027999A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 143 Human Secreted Proteins
; FILE REFERENCE: P5300P1
; CURRENT APPLICATION NUMBER: US/09/986,480
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: PCT/US00/12788
; PRIOR FILING DATE: 2000-05-11
; PRIOR APPLICATION NUMBER: US 60/134,068
; PRIOR FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 456
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 50
; TYPE: DNA
; LENGTH: 3337
; ORGANISM: Homo sapiens
US-09-986-480-50

Query Match      83.3%; Score 15; DB 10; Length 3337;
Best Local Similarity 88.2%; Pred. No. 2.3e+02;
Matches 15; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 TGTGGGAGTTATCGAG 17
Db      2999 TGTGGGAGTTCTCGG 2983

RESULT 6
US-09-895-298-51/c
; Sequence 51, Application US/09895298
; Publication No. US20030078405A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 47 Human Secreted Proteins
; FILE REFERENCE: P2035P1
; CURRENT APPLICATION NUMBER: US/09/895,298
; CURRENT FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: 09/591,16
; PRIOR FILING DATE: 2000-08-09
; PRIOR APPLICATION NUMBER: PCT/US99/29950
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: 60/113,006
; PRIOR FILING DATE: 1998-12-18
; PRIOR APPLICATION NUMBER: 60/112,809
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 51
; LENGTH: 3337
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-895-298-51

Query Match      83.3%; Score 15; DB 10; Length 3337;
Best Local Similarity 88.2%; Pred. No. 2.3e+02;
Matches 15; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 TGTGGGAGTTATCGAG 17
Db      2999 TGTGGGAGTTCTCGG 2983

RESULT 7
US-09-918-995-24221
; Sequence 24221, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
```

```
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 24221
; LENGTH: 468
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(468)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-24221

Query Match      82.2%; Score 14.8; DB 10; Length 468;
Best Local Similarity 88.9%; Pred. No. 2.7e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGTGGGAGTTATCGAGC 18
Db      315 TGTGGGAGTTATCAAC 332

RESULT 8
US-09-878-178-963
; Sequence 963, Application US/09878178
; Patent No. US20020177552A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 21021.527
; CURRENT APPLICATION NUMBER: US/09/878,178
; CURRENT FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 2237
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 963
; LENGTH: 539
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-878-178-963

Query Match      82.2%; Score 14.8; DB 9; Length 539;
Best Local Similarity 88.9%; Pred. No. 2.7e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGTGGGAGTTATCGAGC 18
Db      62 TGTGGAGATTATAGAGC 79

RESULT 9
US-10-046-935-963
; Sequence 963, Application US/10046935
; Publication No. US20020156011A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; APPLICANT: Wang, Aljun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 21021.527C1
; CURRENT APPLICATION NUMBER: US/10/046,935
; CURRENT FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 2239
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 963
; LENGTH: 539
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-046-935-963

Query Match      82.2%; Score 14.8; DB 14; Length 539;
Best Local Similarity 88.9%; Pred. No. 2.7e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGTGGGAGTTATCGAGC 18
   |||||
Db 62 TGTGGAGATTATAGAC 79

RESULT 10
US-10-146-502-963
; Sequence 963, Application US/10146502
; Publication No. US20030069180A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; APPLICANT: Wang, Aijun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.527C2
; CURRENT APPLICATION NUMBER: US/10/146,502
; CURRENT FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2241
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 963
; LENGTH: 539
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-146-502-963

Query Match      82.2%; Score 14.8; DB 15; Length 539;
Best Local Similarity 88.9%; Pred. No. 2.7e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGTGGGAGTTATCGAGC 18
   |||||
Db 62 TGTGGAGATTATAGAC 79

RESULT 11
US-09-738-626-2293
; Sequence 2293, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
```

```
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 2293
; LENGTH: 738
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-2293

Query Match      82.2%; Score 14.8; DB 9; Length 738;
Best Local Similarity 88.9%; Pred. No. 2.8e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGTGGGAGTTATCGAGC 18
   |||||
Db 452 TGTGGGAAGTTCTCGAGC 469

RESULT 12
US-10-027-632-257422
; Sequence 257422, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 257422
; LENGTH: 1058
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-257422

Query Match      82.2%; Score 14.8; DB 13; Length 1058;
Best Local Similarity 88.9%; Pred. No. 2.8e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGTGGGAGTTATCGAGC 18
   |||||
Db 307 TGTGGGAGTTCTCGAGC 324

RESULT 13
US-10-027-632-257423
; Sequence 257423, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
```

```
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 257423
; LENGTH: 1058
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-257423
```

```
Query Match      82.2%; Score 14.8; DB 13; Length 1058;
Best Local Similarity 88.9%; Pred. No. 2.8e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      1  TGTGGGGAGTTATCGAGC 18
          |||||
DB      307  TGTGGGGAGTTCTGGAGC 324
```

```
RESULT 14
US-10-027-632-257422
; Sequence 257422, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 257422
; LENGTH: 1058
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-257422
```

```
Query Match      82.2%; Score 14.8; DB 16; Length 1058;
Best Local Similarity 88.9%; Pred. No. 2.8e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      1  TGTGGGGAGTTATCGAGC 18
          |||||
DB      307  TGTGGGGAGTTCTGGAGC 324
```

```
RESULT 15
US-10-027-632-257423
; Sequence 257423, Application US/10027632
```

```
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 257423
; LENGTH: 1058
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-257423
```

```
Query Match      82.2%; Score 14.8; DB 16; Length 1058;
Best Local Similarity 88.9%; Pred. No. 2.8e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      1  TGTGGGGAGTTATCGAGC 18
          |||||
DB      307  TGTGGGGAGTTCTGGAGC 324
```

```
Search completed: April 24, 2004, 23:47:33
Job time : 141.111 secs
```

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	16.4	91.1	612	13	BQ794028	BQ794028 EST 2965
2	16.4	91.1	829	14	CF513180	CF513180 Cabud0003
3	16.4	91.1	900	28	CC138273	CC138273 NDL 7707.
4	16.4	91.1	952	14	CA964431	CA964431 CCLX04a14

Seq primer: T7.
 Location/Qualifiers
 1..612
 /organism="Vitis vinifera"
 /mol_type="mRNA"
 /cultiivar="Shiraz"
 /db_xref="taxon:29760"
 /clone="TW251H04"
 /dev_stage="veraison stage"
 /clone_lib="Veraison Grape berries Lambda Zap II Library"
 /note="Organ: Fruit; Vector: Lambda Zap II; Site_1: Eco RI; Site_2: XhoI; Oriented library, construction described in Generation of ESTs from grape berry (skin, pulp or seeds) at various developmental stages by Terrier N., Ageorges A., Abbal P., Romieu C. in J. Plant Physiol. 158 (12): 1575-83 2001"

ORIGIN
 Query Match 91.1%; Score 16.4; DB 13; Length 612;
 Best Local Similarity 94.4%; Pred. No. 1.2e+03;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGTGGGGAGTTATCGAGC 18
 |||||
 DB 517 TGTGGGGAGTTATCGAGC 534
 |||||

RESULT 2
 CF512180
 LOCUS
 DEFINITION
 CABud0003 IF_D08 Vitis vinifera cv. cabernet sauvignon (Clone 8)
 Bud - CABUD Vitis vinifera cDNA clone CABud0003_IF_D08 5', mRNA sequence.

ACCESSION
 CF512180
 VERSION
 CF512180.1 GI:34543948
 KEYWORDS
 EST.
 SOURCE
 Vitis vinifera
 ORGANISM
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; Vitaceae; Vitis.
 1 (bases 1 to 829)
 REFERENCE
 Goes da Silva, F., Iandolino, A., Lim, H., Baek, J., Jones, K. and Cook, D.
 Expressed sequence tags from Vitis vinifera 'Cabernet sauvignon' berries at various developmental stages
 Unpublished (2003)
 CONTACT: Douglas Cook, PhD
 CAES Genome Facility
 UC Davis, Plant Pathology
 One Shields Ave, Davis, CA 95616, USA
 Tel: 530 754 6561
 Fax: 530 754 6617
 Email: drcook@ucdavis.edu
 Seq primer: ACGGTACCGACATATGCC.
 Location/Qualifiers
 1..829
 /organism="Vitis vinifera"
 /mol_type="mRNA"
 /cultiivar="Cabernet Sauvignon (Clone 8)"
 /db_xref="taxon:29760"
 /clone="CABud0003_IF_D08"
 /sex="Hermaphrodite"
 /dev_stage="Pre-bloom (10-11 days before bloom)"
 /lab_host="DH5alpha"
 /clone_lib="Vitis vinifera cv. cabernet sauvignon (Clone 8) Bud - CABUD"
 /note="Organ: Bud; Vector: pDNR; Site_1: SfiI; Site_2: SfiI; CABUD is a cDNA library of Vitis vinifera cv. 'Cabernet Sauvignon' clone 8 dissected buds. Samples were collected May 13, 2002 from pre-bloom plants (10-11 days before bloom), pre-veraison. Sampled vines were located at the University of California, Davis, Experimental

FEATURES
 source

Vineyard. cDNAs were made by oligo-dT priming and directionally cloned. 5' and 3' adaptors were used in cloning as follows:
 5'-AAGCAGTGTATTCACGACAGTGGCCATTACGGCCGG-3' and
 5'-ATTCTAGGCGGAGCGGCCACATG-dT(30)NN-3'. Library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.5-3 kb size fraction."

ORIGIN
 Query Match 91.1%; Score 16.4; DB 14; Length 829;
 Best Local Similarity 94.4%; Pred. No. 1.2e+03;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGTGGGGAGTTATCGAGC 18
 |||||
 DB 107 TGTGGGGAGTTATCGAGC 124
 |||||

RESULT 3
 CC138273
 LOCUS
 DEFINITION
 NDL.7707.77 Notre Dame Liverpool Aedes aegypti genomic clone
 NDL.7707, genomic survey sequence.

ACCESSION
 CC138273
 VERSION
 CC138273.1 GI:30007328
 KEYWORDS
 GSS.
 SOURCE
 Aedes aegypti (yellow fever mosquito)
 ORGANISM
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Aedes; Stegomyia.
 1 (bases 1 to 900)
 REFERENCE
 Loftus B., Shetty, J., Knudson, D. and Severson, D.
 BAC end sequencing of Aedes aegypti
 Unpublished (2003)
 JOURNAL
 Other GSSs: NDL.7707.SP6
 COMMENT
 Contact: Brendan Loftus
 Department of Eukaryotic Genomics
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-838-3543
 Fax: 301-838-0208
 Email: enta@tigr.org
 Library was provided by David Severson
 Seq primer: T7
 Class: BAC ends.
 Location/Qualifiers
 1..900
 /organism="Aedes aegypti"
 /mol_type="genomic DNA"
 /strain="liverpool"
 /db_xref="taxon:7159"
 /clone="NDL.7707"
 /clone_lib="Notre Dame Liverpool"
 /note="Vector: pCEBAC1; Site_1: Hind III; The library was prepared from whole body tissue of newly hatched L1 larvae by David Severson at the University of Notre Dame and Hongbin Zhang"

FEATURES
 source

ORIGIN
 Query Match 91.1%; Score 16.4; DB 28; Length 900;
 Best Local Similarity 94.4%; Pred. No. 1.2e+03;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGTGGGGAGTTATCGAGC 18
 |||||
 DB 852 TGTGGGGAGTTATCGAGC 869
 |||||

RESULT 4
 CA964431
 LOCUS
 DEFINITION
 CcLX04a14120f1 Carp mixed tissue library 1 Cyprinus carpio cDNA

clone 14120 5', mRNA sequence.
 CA964431
 VERSION CA964431.1 GI:27490988
 KEYWORDS EST.
 SOURCE
 ORGANISM *Cyprinus carpio* (common carp)
Cyprinus carpio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; *Cyprinus*.
 1 (bases 1 to 952)
 Gracey, A.Y., Fraser, E., Li, W. and Cossins, A.R.
 Microarray and EST analysis of the carp (*Cyprinus carpio*)
 transcriptome during environmental stress
 Unpublished (2003)
 CONTACT: Andrew R. Cossins
 Laboratory for Environmental Gene Regulation
 University of Liverpool
 School of Biological Sciences, The Biosciences Building, Crown
 Street, Liverpool, United Kingdom, L69 7ZB
 Tel: +44(0)151-795-4510
 Fax: +44(0)151-795-4431
 Email: cossins@liv.ac.uk
 Vector has been trimmed from this EST.
 Plate: 14 row: 1 column: 20
 Seq primer: Triplex 5' LD (5'-CTCGGAGCGCGCCATTGTGTGGT-3')
 High quality sequence start: 32
 High quality sequence stop: 685.
 Location/Qualifiers
 1..952
 /organism="Cyprinus carpio"
 /mol_type="mRNA"
 /db_xref="taxon.7962"
 /clone="14120"
 /sex="Male & female"
 /tissue_type="Skeletal white muscle, cardiac muscle,
 kidney, brain, gill, intestinal mucosa"
 /dev_stage="Adult"
 /lab_host="E.coli Electromax DH10B"
 /clone_lib="Carp mixed tissue library 1"
 /notes="Vector: Triplex2; Site 1: Sfil GGCCATTACGGCC;
 Site 2: Sfil GGCGCTCGGCC; Normalized cDNA library
 prepared from mixed tissues of warm, cold and hypoxia
 challenged animals"
 ORIGIN
 Query Match 91.1%; Score 16.4; DB 14; Length 952;
 Best Local Similarity 94.4%; Pred. No. 1.2e+03;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 TGTGGGAGTTATCGAGC 18
 |||
 Db 49 TGTGGGAGTTATCGAGC 66
 |||
 RESULT 5
 BU508464/c
 LOCUS
 DEFINITION AGENCOURT 10094654 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:6502554
 5', mRNA sequence.
 ACCESSION BU508464
 VERSION BU508464.1 GI:22814697
 KEYWORDS EST.
 SOURCE
 ORGANISM Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1039)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM14058 row: m column: 19
 High quality sequence stop: 387.
 Location/Qualifiers
 1..1039
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon.9606"
 /clone="IMAGE:6502554"
 /tissue_type="leiomyosarcoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC 71"
 /note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 2.1 kb."
 ORIGIN
 Query Match 91.1%; Score 16.4; DB 13; Length 1039;
 Best Local Similarity 94.4%; Pred. No. 1.2e+03;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 TGTGGGAGTTATCGAGC 18
 |||
 Db 1020 TGTGGGAGTTATCGAGC 1003
 |||
 RESULT 6
 BH575853/c
 LOCUS
 DEFINITION BOGDJ69TF BOGD Brassica oleracea genomic clone BOGDJ69, genomic
 survey sequence.
 ACCESSION BH575853
 VERSION BH575853.1 GI:17828134
 KEYWORDS GSS.
 SOURCE Brassica oleracea
 ORGANISM Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 394)
 Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
 Whole genome shotgun sequencing of Brassica oleracea
 Unpublished (2001)
 Other_GSSs: BOGDJ69TR
 CONTACT: Chris Town
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TF
 Class: sheared ends.
 Location/Qualifiers
 1..394
 /organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /strain="Tol000DH3"
 /db_xref="taxon:3712"
 /clone="BOGDJ69"
 /note="Vector: pHOS1; Site 1: BstXI; 2-3 kb sheared
 genomic DNA inserted into pHOS1 using BstXI linkers"
 ORIGIN
 Query Match 88.9%; Score 16; DB 28; Length 394;
 Best Local Similarity 100.0%; Pred. No. 1.8e+03;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY      1  TGTGGGAGTTATCGA 16
Db      177 TGTGGGAGTTATCGA 162

RESULT 7
BZ473722
LOCUS   BZ473722 BO 1.6 2 KB tot Brassica oleracea genomic clone BONLX22,
DEFINITION      830 bp DNA linear GSS 13-DEC-2002
genomic survey sequence.
ACCESSION      BZ473722
VERSION        BZ473722.1 GI:26773982
KEYWORDS       GSS.
SOURCE         Brassica oleracea
ORGANISM       Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE      1 (bases 1 to 830)
AUTHORS        Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
TITLE          Whole genome shotgun sequencing of Brassica oleracea
JOURNAL        Unpublished (2001)
COMMENT        Other GSSs: BONLX22TR
Contact: Chris Town
TIGR Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: Sheared ends.
FEATURES             Location/Qualifiers
     source           1..830
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="T01000DH3"
                     /db_xref="taxon:3712"
                     /clone="BONLX22"
                     /clone_lib="BO_1.6_2_KB_tot"
                     /note="Vector: pHOSt; Site 1: BstXI; 1.6-2 kb sheared
total DNA inserted into pHOSt using BstXI linkers"

ORIGIN
Query Match      88.9%; Score 16; DB 28; Length 830;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  TGTGGGAGTTATCGA 16
Db      255 TGTGGGAGTTATCGA 270

RESULT 8
CG744540/c
LOCUS   CG744540 1230 bp DNA linear GSS 24-OCT-2003
DEFINITION      P037-1-F08.ya Ppa EcoRI BAC library Pristionchus pacificus genomic,
genomic survey sequence.
ACCESSION      CG744540
VERSION        CG744540.1 GI:37965408
KEYWORDS       GSS.
SOURCE         Pristionchus pacificus
ORGANISM       Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Neodiplogasteridae; Pristionchus.
REFERENCE      1 (bases 1 to 1230)
AUTHORS        Srinivasan, J., Sins, W., Jesse, T., Wiggers-Perebolte, L., Jansen, K.,
Buitjer, J., van der Meulen, M. and Sommer, R. J.
TITLE          An integrated physical and genetic map of the nematode Pristionchus
pacificus
JOURNAL        Mol. Genet. Genomics 269 (5), 715-722 (2003)
MEDLINE        22835951

```

```

12884007
Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39 Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.bommer@tuebingen.mpg.de
Class: BAC ends.
FEATURES             Location/Qualifiers
     source           1..1230
                     /organism="Pristionchus pacificus"
                     /mol_type="genomic DNA"
                     /strain="California"
                     /db_xref="taxon:54126"
                     /clone_lib="Ppa EcoRI BAC Library"
                     /note="The library was generated by a partial digest of
the genomic DNA with EcoRI and cloning into the BAC
vector."

ORIGIN
Query Match      88.9%; Score 16; DB 29; Length 1230;
Best Local Similarity 100.0%; Pred. No. 2e+03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3  TGGGGAGTTATCGAGC 18
Db      301 TGGGGAGTTATCGAGC 286

RESULT 9
BG072494
LOCUS   BG072494 297 bp mRNA linear EST 26-JAN-2001
DEFINITION      H3111C05-3 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
H3111C05 3', mRNA sequence.
ACCESSION      BG072494
VERSION        BG072494.1 GI:12555063
KEYWORDS       EST.
SOURCE         Mus musculus (house mouse)
ORGANISM       Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 297)
REFERENCE      Kargul, G.J., Dudekula, D.B., Qian, Y., Lim, M.K., Jaradat, S.A.,
Tanaka, T.S., Carter, M.G. and Ko, M.S.H.
AUTHORS        Verification and initial annotation of NIA mouse 15K cDNA clone set
Unpublished (2001)
TITLE          Other ESTs: H3111C05-5
JOURNAL        Contact: George J. Kargul
COMMENT        Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@gsun.grc.nia.nih.gov
This clone set has been freely distributed to the community. Please
visit http://lgsun.grc.nia.nih.gov/cDNA/15k.html for details.
Plate: H3111 row: C column: 05
Seq primer: -21M3 Forward
High quality sequence stop: 297
POLYA=Yes.
FEATURES             Location/Qualifiers
     source           1..297
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="niaEST:H3111C05-3"
                     /db_xref="taxon:10090"
                     /clone="H3111C05"
                     /dev_stage="Clones arrayed from a variety of cDNA libraries"
                     /lab_host="DH108"
                     /clone_lib="NIA Mouse 15K cDNA Clone Set"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This

```


clone is among a rearranged set of 15,247 clones from 11 embryo cDNA libraries (including preimplantation stage embryos from unfertilized egg to blastocyst, embryonic part of E7.5 embryos, extraembryonic part of E7.5 embryos, and E12.5 female mesonephros/gonad) and one newborn ovary cDNA library. Average insert size 1.5 kb. All source libraries are cloned unidirectionally with Oligo(dT)-Not primers. References include: (1) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray. 2000, Proc. Natl. Acad. Sci. U S A. 97: 9127-9132; (2) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development, 2000, Development, 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."

ORIGIN

Query Match 85.6%; Score 15.4; DB 10; Length 297;
Best Local Similarity 94.1%; Pred. No. 3.4e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GTGGGAGTTATCGAGC 18

|||||

Db 234 GTGGGAGTTATCGAGC 250

RESULT 10

AA034988

LOCUS

DEFINITION AA034988 322 bp mRNA linear EST 10-MAY-1997
IMAGE:471608 5', mRNA sequence.

ACCESSION AA034988

VERSION AA034988.1 GI:1506951

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

REFERENCE 1 (bases 1 to 322)
Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B., Chisoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,P., Thierry-Veg,J., Trevaskis,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.

REFERENCE

AUTHORS

TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)
MEDLINE 9704478
PUBMED 889549

COMMENT

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1181 Std Error: 0.00
Seq primer: -28M13 rev2 from Amerham
High quality sequence stop: 255.

FEATURES

source

1..322
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:3757610"
/db_xref="taxon:9606"
/clone="IMAGE:471608"
/sex="female"
/dev_stage="adult"

/lab_host="DH10B"
/clone_lib="Soares pregnant uterus NBHPU"
/note="Organ: uterus; Vector: pTV3-Pac; Site 1: Not I;
Site 2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5',
AATGGAGATTCGGCGCGCTTTT TTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pTV3 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."

ORIGIN

Query Match 85.6%; Score 15.4; DB 9; Length 322;
Best Local Similarity 88.9%; Pred. No. 3.4e+03;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGTGGGAGTTATCGAGC 18

|||||

Db 254 TGTGGGAGTTATCGAGC 271

RESULT 11

AV659377/c

LOCUS

DEFINITION AV659377 GLC Homo sapiens cDNA clone G1CFWD12 3', mRNA sequence. EST 16-JAN-2002

ACCESSION AV659377

VERSION AV659377.1 GI:9880391

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

REFERENCE 1 (bases 1 to 353)
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
Hu,G., Gu,J., Chen,Z. and Han,Z.

TITLE

Insight into hepatocellular carcinogenesis at transcriptome level
by comparing gene expression profiles of hepatocellular carcinoma
with those of corresponding noncancerous liver

JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922

Email: hanzg@chgc.sh.cn

This clone is available at CHGC in Shanghai.

Location/Qualifiers

FEATURES

source

1..353
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="G1CFWD12"
/tissue_type="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/clone_lib="GLC"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

ORIGIN

Query Match 85.6%; Score 15.4; DB 9; Length 353;
Best Local Similarity 94.1%; Pred. No. 3.4e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGTGGGAGTTATCGAG 17

|||||

Db 212 TGTGGGAGTTCTCGAG 196

RESULT 12
 AV659577/c
 LOCUS AV659577 364 bp mRNA linear EST 16-JAN-2002
 DEFINITION AV659577 GLC Homo sapiens cDNA clone GLCFY11 3', mRNA sequence.
 ACCESSION AV659577
 VERSION AV659577.1 GI:9880591
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 364)
 AUTHORS Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
 Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
 Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
 Hu,G., Gu,J., Chen,Z. and Han,Z.
 TITLE Insight into hepatocellular carcinogenesis at transcriptome level
 by comparing gene expression profiles of hepatocellular carcinoma
 with those of corresponding noncancerous liver
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
 MEDLINE 21625106
 PUBMED 11752456
 COMMENT Contact: Zeguang Han
 Chinese National Human Genome Center at Shanghai
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
 201203, P. R. China
 Tel: 86-21-50801919(ex.45)
 Fax: 86-21-50801922
 Email: hanzg@chgc.sh.cn
 This clone is available at CHGC in Shanghai.
 FEATURES
 source
 Location/Qualifiers
 1..364
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="GLCFY11"
 /tissue_type="corresponding non cancerous liver tissue"
 /dev_stage="Adult"
 /lab_host="SOLR"
 /clone_lib="GLC"
 /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
 XhoI"

Query Match 85.6%; Score 15.4; DB 9; Length 364;
 Best Local Similarity 94.1%; Pred. No. 3.4e+03;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 TGTGGGAGTTATCGAG 17
 |||||
 Db 212 TGTGGGAGTTCTCGAG 196
 |||||

RESULT 13
 AV659453/c
 LOCUS AV659453 368 bp mRNA linear EST 16-JAN-2002
 DEFINITION AV659453 GLC Homo sapiens cDNA clone GLCFXC06 3', mRNA sequence.
 ACCESSION AV659453
 VERSION AV659453.1 GI:9880467
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 368)
 AUTHORS Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
 Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
 Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
 Hu,G., Gu,J., Chen,Z. and Han,Z.
 TITLE Insight into hepatocellular carcinogenesis at transcriptome level
 by comparing gene expression profiles of hepatocellular carcinoma
 with those of corresponding noncancerous liver

with those of corresponding noncancerous liver
 Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
 MEDLINE 21625106
 PUBMED 11752456
 COMMENT Contact: Zeguang Han
 Chinese National Human Genome Center at Shanghai
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
 201203, P. R. China
 Tel: 86-21-50801919(ex.45)
 Fax: 86-21-50801922
 Email: hanzg@chgc.sh.cn
 This clone is available at CHGC in Shanghai.
 FEATURES
 source
 Location/Qualifiers
 1..368
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="GLCFXC06"
 /tissue_type="corresponding non cancerous liver tissue"
 /dev_stage="Adult"
 /lab_host="SOLR"
 /clone_lib="GLC"
 /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
 XhoI"

Query Match 85.6%; Score 15.4; DB 9; Length 368;
 Best Local Similarity 94.1%; Pred. No. 3.5e+03;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 TGTGGGAGTTATCGAG 17
 |||||
 Db 212 TGTGGGAGTTCTCGAG 196
 |||||

RESULT 14
 AV659465/c
 LOCUS AV659465 374 bp mRNA linear EST 16-JAN-2002
 DEFINITION AV659465 GLC Homo sapiens cDNA clone GLCFXD07 3', mRNA sequence.
 ACCESSION AV659465
 VERSION AV659465.1 GI:9880479
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 374)
 AUTHORS Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
 Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
 Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
 Hu,G., Gu,J., Chen,Z. and Han,Z.
 TITLE Insight into hepatocellular carcinogenesis at transcriptome level
 by comparing gene expression profiles of hepatocellular carcinoma
 with those of corresponding noncancerous liver
 Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
 MEDLINE 21625106
 PUBMED 11752456
 COMMENT Contact: Zeguang Han
 Chinese National Human Genome Center at Shanghai
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
 201203, P. R. China
 Tel: 86-21-50801919(ex.45)
 Fax: 86-21-50801922
 Email: hanzg@chgc.sh.cn
 This clone is available at CHGC in Shanghai.
 FEATURES
 source
 Location/Qualifiers
 1..374
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="GLCFXD07"
 /tissue_type="corresponding non cancerous liver tissue"
 /dev_stage="Adult"

```

/lab_host="SOLR"
/clone_lib="GLC"
/notes=vector: pbluescript sk(-); Site_1: EcoRI; Site_2:
XhoI

ORIGIN
Query Match      85.6%; Score 15.4; DB 9; Length 374;
Best Local Similarity 94.1%; Pred. No. 3.5e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGTGGGGAGTTATCGAG 17
    |||||
Db 212 TGTGGGGAGTTATCGAG 196

RESULT 15
BY479676/c
LOCUS      526 bp      mRNA      linear      EST 02-JAN-2003
DEFINITION BY479676 RIKEN full-length enriched, B16 F10Y cells Mus musculus
cDNA clone G370120F12 3', mRNA sequence.
ACCESSION  BY479676
VERSION    BY479676.1 GI:26814055
KEYWORDS   EST.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 526)
Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamataka,I.,
Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Bruscia,V.,
Ciothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
Kawaji,H., Kawasawa,Y., Kedierski,R.M., King,B.L., Konagaya,A.,
Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
Viarado,R., Wagner,L., Wahlstedt,C., Wang,Y., Watanabe,Y.,
Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
Rogers,J., Birney,E. and Hayashizaki,Y.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
JOURNAL    22354683
MEDLINE    22354683
PUBMED     12466851
COMMENT    Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-resgsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,
Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,
Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,
Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,

```

```

Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and
Hayashizaki,Y. Direct Submission
Computational Analysis of Full-length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Cells were provided by Drs. William J Pavan, Stacie Loftus, and
Denise Larson (Division of Intramural Research Genetic Disease
Research Branch National Human Genome Research Institute, National
Institutes of Health (NIH) Building: 49, Room 4A82 49 Convent Drive
MSC 4472 Bethesda, Maryland U.S.A) whose assistance we gratefully
acknowledge.
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
FEATURES             Location/Qualifiers
     source           1..526
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="G370120F12"
                     /cell_type="B16 F10Y cells"
                     /clone_lib="RIKEN full-length enriched, B16 F10Y cells"

ORIGIN
Query Match      85.6%; Score 15.4; DB 13; Length 526;
Best Local Similarity 94.1%; Pred. No. 3.6e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGTGGGGAGTTATCGAG 17
    |||||
Db 94 TGTGGGGAGTTATCGAG 78

Search completed: April 24, 2004, 21:10:55
Job time : 1406.11 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2004, 17:02:49 ; Search time 562.494 Seconds
(without alignments)
1464.047 Million cell updates/sec

Title: US-10-084-555A-118

Perfect score: 19

Sequence: 1 gccttcgcgaataatcg 19

Scoring table:

IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.pat.*
- 6: gb.ph.*
- 7: gb.pl.*
- 8: gb.pr.*
- 9: gb.ro.*
- 10: gb.sts.*
- 11: gb.un.*
- 12: gb.un.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.mam.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	18	94.7	62730	2	AC020031	Drosophila
C 2	18	94.7	170498	3	AC008094	Drosophila
C 3	18	94.7	244757	3	AE003678	Drosophila
C 4	17.4	91.6	126447	8	OST00189	AL662985 Oryza sat
C 5	17	89.5	2669	1	AB013090	AB013090 Clostridi
C 6	17	89.5	2669	1	AB013091	AB013091 Clostridi
C 7	17	89.5	2751	1	AB013089	AB013089 Clostridi
C 8	17	89.5	6159	8	AB079543	AB079543 Schizosac
C 9	17	89.5	43702	8	SPAC15A10	Z97208 S.pombe chr
C 10	17	89.5	132930	2	AC079046	AB079046 Homo sapi
C 11	17	89.5	157690	8	CNS08CA3	AL772425 Oryza sat
C 12	16.4	86.3	386	6	AX887299	AX887299 Sequence
C 13	16.4	86.3	386	6	BD026909	BD026909 Sequence
C 14	16.4	86.3	20097	1	AE008752	AE008752 Salmonell
C 15	16.4	86.3	42703	2	AC145319	AC145319 Oryza sat
C 16	16.4	86.3	63032	9	AC117419	AC117419 Homo sapi
C 17	16.4	86.3	100814	2	AP003929	AP003929 Oryza sat
C 18	16.4	86.3	144007	2	AC141743	AC141743 Apis mell
C 19	16.4	86.3	151223	2	AC141710	AC141710 Apis mell
C 20	16.4	86.3	152365	8	AP005167	AP005167 Oryza sat
C 21	16.4	86.3	158700	2	AC027643	AC027643 Homo sapi
C 22	16.4	86.3	175292	2	AC022632	AC022632 Homo sapi
C 23	16.4	86.3	180353	9	AC079326	AC079326 Homo sapi
C 24	16.4	86.3	180708	9	AL590725	AL590725 Human DNA
C 25	16.4	86.3	184654	2	AC134045	AC134045 Oryza sat
C 26	16.4	86.3	254050	1	AL627269	AL627269 Salmonell
C 27	16.4	86.3	300029	1	AE016839	AE016839 Salmonell
C 28	16.4	86.3	343116	1	AP003003	AP003003 Mesorhizo
C 29	16	84.2	4487	6	E00825	E00825 DNA fragmen
C 30	16	84.2	4487	6	E01599	E01599 DNA sequenc
C 31	16	84.2	5482	1	BACGNT	J02584 B.subtilis
C 32	16	84.2	10164	1	AE011812	AE011812 Xanthomon
C 33	16	84.2	13719	3	AC006654	AC006654 Caenorhab
C 34	16	84.2	36448	1	AB005554	AB005554 Bacillus
C 35	16	84.2	38475	3	AF078786	AF078786 Caenorhab
C 36	16	84.2	58829	2	AC087520	AC087520 Homo sapi
C 37	16	84.2	85397	3	AC026301	AC026301 Caenorhab
C 38	16	84.2	128679	8	CNS08CDB	BX000345 Oryza sat
C 39	16	84.2	131530	8	AC122143	AC122143 Oryza sat
C 40	16	84.2	147341	8	AC122143	AC122143 Oryza sat
C 41	16	84.2	147720	2	AC141845	AC141845 Apis mell
C 42	16	84.2	165169	2	AC102380	AC102380 Mus muscu
C 43	16	84.2	203901	1	BSUB0021	Z99124 Bacillus su
C 44	16	84.2	224802	2	AC006787	AC006787 Caenorhab
C 45	16	84.2	298216	2	AC006875	AC006875 Caenorhab

ALIGNMENTS

RESULT 1
AC020031/c
LOCUS AC020031 62730 bp DNA linear HTG 03-JAN-2000
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION AC020031
VERSION AC020031.1 GI:6664866
KEYWORDS HTG; HTGS PHASE2.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE Adams, M. and Venter, J.C.
AUTHORS Direct Submission
TITLE 1 (bases 1 to 62730)

JOURNAL Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA

COMMENT This sequence was identified as CDM10211802 by the submitter. For more information on this record e-mail to fly@celera.com.

* NOTE: This is a 'working draft' sequence.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

FEATURES
source
1..62730
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"

ORIGIN
Query Match 94.7%; Score 18; DB 2; Length 62730;
Best Local Similarity 100.0%; Pred. NO. 1.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCTTCGCGAAAAAATC 18
|||||

Db 54065 GCCTTCGCGAAAAAATC 54048

RESULT 2
AC008094 170498 bp DNA linear INV 21-FEB-2001
LOCUS Drosophila melanogaster, chromosome 3R, region 84F-84F, BAC clone
DEFINITION BACR45A07, complete sequence.

ACCESSION AC008094
VERSION AC008094.5 GI:13027527
KEYWORDS HTG.

SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster

REFERENCE
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 170498)

AUTHORS
Celniker S.E., Adams M.D., Kronmiller B., Tyler D., Wan K.H.,
Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C.,
Rogers Y., An H., Baldwin D., Banazon J., Beeson K.Y., Busam D.A.,
Carlson J.W., Center A., Champs M., Davenport L.B., Dietz S.M.,
Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
Farrera S., Frise B., Galle R.F., Garg N.S., George R.A.,
Gonzalez M., Hock J., Hoskins R.A., Hostin D., Howland T.J.,
Ibgwegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoc J.,
Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
Stapleton M., Strong R., Svirskaas R., Tector C., Williams S.M.,
Zaveri J.S., Smith H.O., Rubin G.M. and Venter J.C.

TITLE Sequencing of Drosophila chromosome 3R, region 84F-84F
Unpublished

REFERENCE
2 (bases 1 to 170498)
Celniker S.E., Agbayani A., Arcaina T.T., Baxter E., Blazej R.G.,
Buenhoff C., Champs M., Chavez C., Chew M., Cieiolka L.,
Doyle C.M., Farfan D.E., Galle R., George R.A., Harris N.L.,
Hoskins R.A., Houston K.A., Hummasti S.R., Karra K., Kearney L.,
Kim E., Lee B., Lewis S., Li P., Lomotan M.A., Mazda P.,
Moshrefi A.R., Moshrefi M., Nixon K., Pacleb J.M., Park S.,
Pfeiffer B., Poon L., Sequeira A., Sethi H., Snir E.,
Svirskaas R.R., Wan K.H., Weinburg T., Zhang R., Zietan L.L. and
Rubin G.M.

TITLE Direct Submission
JOURNAL Submitted (22-JUL-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT On Feb 21, 2001 this sequence version replaced gi:6957909.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.

For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (<http://www.fruitfly.org/sequence/>) or send email
to bdgp@fruitfly.berkeley.edu.

Location/Qualifiers
1..170498
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="3R"
/map="84F-84F"
/clone="BACR45A07 (D827)"
/clone_lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
pBACE3.6)"

FEATURES
source
1..170498
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="3R"
/map="84F-84F"
/clone="BACR45A07 (D827)"
/clone_lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
pBACE3.6)"

ORIGIN
Query Match 94.7%; Score 18; DB 3; Length 170498;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCTTCGCGAAAAAATC 18
|||||

Db 133844 GCCTTCGCGAAAAAATC 133861

RESULT 3
AE003678 244757 bp DNA linear INV 14-FEB-2003
LOCUS Drosophila melanogaster chromosome 3R, section 16 of 118 of the
DEFINITION complete sequence.

ACCESSION AE003678 AE002708 AE014297
VERSION AE003678.3 GI:23175896
KEYWORDS Drosophila melanogaster (fruit fly)
SOURCE Drosophila melanogaster
ORGANISM Drosophila melanogaster

REFERENCE
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 244757)

AUTHORS
Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
Sutton R.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
Brandon R.C., Rogers J.H., Blazej R.G., Champs M., Pfeiffer B.D.,
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabot G.D.,
Abril J.F., Agbayani A., An H.J., Andrews-Frankoch C., Baldwin D.,
Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A.,
Chandra I., Cherry J.M., Cawley S., Dahike C., Davenport L.B.,
Davies P., de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I.,
Dietz S.M., Dodson K., Doup L.E., Downes M., Dugan-Rocha S.,
Dunkov B.C., Dunn P., Durbin K.J., Evangelista C.C., Ferraz C.,
Fieriera S., Fleischmann W., Foster C., Gabrielian A.E., Garg N.S.,
Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H., Gu Z.,
Guan P., Harris M., Harris N.L., Harvey D., Heiman T.J.,
Herrandez R., Hock J., Hostin D., Houston K.A., Howland T.J.,
Wei M.H., Ibgwegwam C., Jalali M., Kalush F., Karpen G.H., Ke Z.,
Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C.,
Kravitz S., Kulp D., Lai Z., Lasko P., Lei Y., Levitsky A.A.,
Li J., Li Z., Liang Y., Lin X., Liu X., Mattei B., McIntosh T.C.,
McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Moberly C.,
Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L.,
Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K.,
Nusskern D.R., Pacleb J.M., Parazolo M., Pittman G.S., Pan S.,
Pollard J., Puri V., Reese M.G., Reinert K., Remington K.,
Saunders R.D., Scheeler F., Shen H., Shue B.C., Siden-Kiamos I.,
Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C.,
Stapleton M., Strong R., Sun E., Svirskaas R., Tector C., Turner R.,
Venter E., Wang A.H., Wang X., Wang Z.Y., Wassarman D.A.,

Weinstock, G.M., Weissenbach, J., Williams, S.M., Woodage, T., Wray, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F., Zaveri, J.S., Zhang, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C.
 The genome sequence of *Drosophila melanogaster*
 Science 287 (5461), 2185-2195 (2000)
 20196006
 10731132
 2 (bases 1 to 244757)
 Celisner, S.E., Adams, M.D., Kronmiller, B., Wan, K.H., Holt, R.A., Evans, C.A., Gocayne, J.D., Anantides, P.G., Brandon, R.C., Rogers, Y., Banzon, J., An, H., Baldwin, D., Banton, J., Beeson, K.Y., Busam, D.A., Carlson, J.W., Center, A., Champs, M., Davenport, L.B., Dietz, S.M., Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Drenek, D., Farfan, D., Ferriz, S., Frise, E., Galle, R.F., Garg, N.S., George, R.A., Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J., Ibeagwam, C., Jalali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A., McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J., Pacleb, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B., Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F., Stapleton, M., Strong, R., Svirskaas, R., Tector, C., Tyler, D., Williams, S.M., Zaveri, J.S., Smith, H.O., Venter, J.C. and Rubin, G.M.
 Sequencing of *Drosophila melanogaster* genome
 Unpublished
 3 (bases 1 to 244757)
 Misra, S., Crosby, M.A., Matthews, B.B., Bayraktaroglu, L., Campbell, K., Hradecky, P., Huang, Y., Kaminker, J.S., Prochownik, S.E., Smith, C.D., Tupy, J.L., Bergman, C.M., Berman, B.P., Carlson, J.W., Celisner, S.E., Clamp, N.E., Drysdale, R.A., Emmert, D., Frise, E., de Grey, A.D.N.J., Harris, N.L., Kronmiller, B., Marshall, B., Millburn, G.H., Richter, J., Russo, S., Searle, S.M.J., Smith, E., Shu, S., Smutniak, F., Whitfield, E.J., Ashburner, M., Gelbart, W.M., Rubin, G.M., Mungall, C.J. and Lewis, S.E.
 Annotation of *Drosophila melanogaster* genome
 Unpublished
 4 (bases 1 to 244757)
 Adams, M.D., Celniker, S.E., Gibbs, R.A., Rubin, G.M. and Venter, C.J.
 Direct Submission
 Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA
 5 (bases 1 to 244757)
 FlyBase
 Direct Submission
 Submitted (06-SEP-2002) University of California Berkeley, 539 Life Sciences Addition, Berkeley, CA 94720, USA
 6 (bases 1 to 244757)
 FlyBase
 Direct Submission
 Submitted (13-FEB-2003) University of California Berkeley, 539 Life Sciences Addition, Berkeley, CA 94720, USA
 Sequence update by submitter
 On Sep 18, 2002 this sequence version replaced gi:10726379.
 Location/Qualifiers
 1. 244757
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"
 /chromosome="3R"
 /notes="genotype: y[1]; cn[1] bw[1] sp[1]; Rh6[1]"
 8444..9229
 /locus tag="TE19336"
 /map="84F3-84F3"
 8444..9229
 /transposon="F-element[1264]"
 /db_xref="FLYBASE:FBst10019336"
 18810..20578
 /locus tag="CG7800"
 /notes="last curated on Fri Jan 11 12:21:03 PST 2002"
 /map="84F4-84F4"
 /db_xref="FLYBASE:FBgn0037552"

18810..20578
 /locus tag="CG7800"
 /product="CG7800-RA"
 /db_xref="FLYBASE:FBgn0037552"
 18933..20494
 /locus tag="CG7800"
 /notes="CG7800 gene product"
 /codon_start=1
 /product="CG7800-PA"
 /protein_id="AAF54196.2"
 /db_xref="GI:10726380"
 /db_xref="FLYBASE:FBgn0037552"
 /translation="MYLIIALLASVALGRPLEDYCYNDYCHLLGRSTFSDKKAT
 KTFEWDSCCKVLMKLNRLTELENCSPDFTWDLNQLPILTSILQRNLLGL
 HDHFSPNPKMILGNNITRLNECFGLAQLMLLSLPGNGIQGLPMDVFNQDPE
 LLDLDSGRNRIETIHENITFGVPEKLEMLNGNPLTAFTSLKSLNELLDSNCG
 PLPLSLGAHTLILNDSGVQRDLILGSHVHKQARKNHITDKLPKSKSVIELDLHSN
 LLATDIPKLLTGMRRLDLSNIGIYAAASDNTSELFILPLMYNLSANRLT
 LRHFDSPKRLTHLDASVNRIVAPKVGIDAEFNLSLHSGNYINNFELTPKPH
 PSLEVALYNKQPKGYKNITFEFNEIGVNVLEKTYQSSNNITPTCKFCIDARDF
 PLSISADNTQITDKVNPUSNTYQWNVNMLVLSLVSFLVNSFLIVKLRLGRN
 QFTQSSVPALIEFNSDESDEIL"
 21001..22999
 /locus tag="CG18249"
 /notes="last curated on Tue Mar 19 11:13:43 PST 2002"
 /map="84F4-84F4"
 /db_xref="FLYBASE:FBgn0037553"
 21001..22999
 /locus tag="CG18249"
 /product="CG18249-RA"
 /db_xref="FLYBASE:FBgn0037553"
 21030..22697
 /locus tag="CG18249"
 /notes="CG18249 gene product"
 /codon_start=1
 /product="CG18249-PA"
 /protein_id="AAF54197.1"
 /db_xref="GI:7298995"
 /db_xref="FLYBASE:FBgn0037553"
 /translation="MHFYKILGICGLTVSDALAVDASGRFNLSACPDSCFSLSD
 RPYASATPALQRLREHLNCSQSITWVLOITPGLRTAVINCAVHISKESLRPV
 ENITSLQWQSTSLGVLQDFNFAVRLIQLSQNFHTVHVAAPQGLSKRLGLQD
 NATAEILSDTDLMEVLHDLDSRLNLTLPQIFAKKKLQTLINGNFPRLIMPDV
 LGSPLNRLDLGHAELVNTLNQNLVLEGSLSLVSINGGFKIOAGNELN
 HLOVGNKSVIEMDLNGLNDTAALLGMWNLQRLDLSKNRIEALPQHSGLDSV
 GTBELLPISLKFILNANOLVRLPPESPLSLSLYLDLSHMLTLDVAILRSLSV
 LKGLYVEGNLTINQKLEHHPDLSELGHDNPWSSGLYKRMFLYFTDRGVHLOAR
 PQRVNLSRVLDLPPNEGQAQKMDPPGTGHPYVTLRDLIAFVTLVVMIL
 LMDYHLIEEGCLRRFRHWRRTLGQNTSTTRNSGRRLDEQDSSEV"
 34339..37638
 /gene="DNapol-iota"
 /locus tag="CG7602"
 /notes="last curated on Tue Mar 19 11:14:07 PST 2002"
 /map="84F4-84F5"
 /db_xref="FLYBASE:FBgn0037554"
 join(34339..34394,35083..36054,36119..37638)
 /gene="DNapol-iota"
 /locus tag="CG7602"
 /product="CG7602-RB"
 /db_xref="FLYBASE:FBgn0037554"
 join(34392..34478,35083..36054,36119..37638)
 /gene="DNapol-iota"
 /locus tag="CG7602"
 /product="CG7602-RA"
 /db_xref="FLYBASE:FBgn0037554"
 join(35098..36054,36119..37375)
 /gene="DNapol-iota"
 /locus tag="CG7602"
 /notes="DNapol-iota gene product from transcript CG7602-RA"
 /codon_start=1
 /product="CG7602-PA"
 /protein_id="AAF54198.1"
 /db_xref="GI:7298996"


```

RQALPFTVTCRCASLAANVVEISGRASPPWMAAGDAVAVIPSYENHVTEVSTCVK
HSSHTDIDRLRLSITTFPHDYDVTILDKIINPLFWC"
/translation="22066..22594"
/genes="OSJNBa0095H06.4"
/complement(join(22066..22594))
/genes="OSJNBa0095H06.4"
/codon_start=1
/protein_id="CA802198.2"
/db_xref="GI:38346779"
/translation="METNOELEALLAGIGEGTSSAMEENDKLFYOEVDRAQSGSK
SQLHVRVKSHTLGMAYLAGAARVNGSPATIGCCSVSGHDVASVPRGKLLQTOPQ
DQLLAHGDYLGLEKRLTSQQLAPAHKVPQAF"
23029..29775
/genes="OSJNBa0095H06.5"
/join(23029..23154,23242..23324,24422..24479,25239..25298,
26266..26395,26481..26707,26886..26978,27055..27126,
28052..28192,28705..28707,29047..29103,29569..29655,
29741..29776)
/genes="OSJNBa0095H06.5"
/codon_start=1
/protein_id="CA802199.2"
/db_xref="GI:38346780"
/translation="MSGGYCAAAVDEPAATELVKRGATLLLDVPEQTLGLVDTVQVF
SVGPKFGIRVSPFGHFLYPCSPNRAHEFAFTVGFLLTHPSEVIVKMAHQSERL
IKLPEEIRKSAVAFHFDPSQLGNDLSDGMDKQLSGLYVQSIVLEPIGGEIT
IANESSMDKAPOTDMERRLMDOLKDGKPAKAPVOSERRGCVYTTIPASIKHNSIG
DELTANLKTCLLESVLAKYQGEDLLGLEQAFAPFAPMGQSLAEAPQWALVSL
LLSCSAFPAIYYQLKHGQHTQDRSEEMGNLFLDEAFSRDIFIRLSKIDFF
AVILEATFVDGDLDSLTRKLSLETTTFGWDLDNNTVNLIIDDEFAPVYVEMDGS"
38637..42270
/genes="OSJNBa0095H06.6"
/join(38637..41550,41777..42270)
/genes="OSJNBa0095H06.6"
/codon_start=1
/protein_id="CA802200.2"
/db_xref="GI:38346781"
/translation="MHFLILLHLLSFSSTASSESAFVITQMLNTTAAAPPDWPSP
SASPCKSHVGDAATSVTSVFQSHLAAPLPFGICAAFLPSLAVSDANTTGG
VPDHLCHRLAIDLNSGLSPGIPASIGNATAMASLANSQSLSGPTPASGLNLA
SLRDLNLRGELPASLGELRLLESIRAGNRDLGGEIPEISFSLNVLVLGALD
TKTSGALPASGLSLOLTSIVTWMLSGIPAEAGCGNLTVNLYENSLSGPIPPS
LGAFLRLKLLNQNSLITGPIPTFGNLSLSVLDLSINAGIPASLGRLPALQDL
MDSNNLTGTPPALANATSLVQLDNTAISGLIPPELGRALAQVFWAQNQLEGS
IPASLAGLANLQDLNHLNLTGAIPEGIFLLNLTKLILLNSGVIPPIEIGKAS
LVRRLGNRLAGTIIPAAVAGMRINFLDGNRLAGVPAELGNSQLQMLDLSNNT
LTCALPESLAGVLEIDVSHNQLTGVDPDAFGRLEALSRLVSGNSLSGAIPAALG
KCNLELLDLSNALSGRIPDELCATDGLDIALNLSRNLGTGPIPARISALSLSVLD
LSYNALDGLAPLAGLDNLVTLNVSNNETGVLPTKLPQLSTCLAGNSGLCTKGG
DVCFSIDASGRVMSADSEEVQMRKLALALVATVAVLGVGLIRARGMGIV
GKGGHGSSESSESGLDLPWQPTFPQKLSFVSVQVNVNLDANIIGKSGSVYR
VGLDTGEVIAVKLWPFSTRNGADKDVAGGRVDSFSAEVRTICRHNTRVFLGC
CNWKTTRLLMYDMYANGSLAVLHERRHGGGGGAQLEWDVRYRIVILGAQGLAYLH
HDCVPTIHRDIKANNILIGLOPEYLTADFLAKLVDGDFGRSSNTVAGSYTAPF
YGYMKITKSDVYVYVVLVLTGKQPIDPTIPDGHVDVWRRRGAADVLPAL
RGSDAEVDMLQVGVALLCVAPSDDRPAMKDVAAMLNEIRLDEDDYANVDLLKS
GAAASAPRAAAATSTSTPPSSSSFSGSSAMYNSSSKAKSPFD"
complement(46428..48083)
/genes="OSJNBa0095H06.7"
/complement(join(46428..46693,46744..47028,47573..47771,
48081..48083))
/genes="OSJNBa0095H06.7"
/codon_start=1
/protein_id="CA802201.2"
/db_xref="GI:38346782"
/translation="MLAARLHAKVCVPRLLLKVDIAKAFDMVSWPFLLEVLQAQFSQ
RWDVSLILASSSVVLVGVGWIKAIIDKRHPFLMTGDSANGQGVSWTKVC
HPRLYSGMGPDLRIKALRVRLWLRQSNPNWDLKASVERSVDMFAASTFTTL
GDASIAFDLLHVVPFRFGRSRTVATGLANNSWVDIRGALTVPVISQFLVWDVLTPI
QLSPGEDRLVWCWTDQCYSVRLAYQAFELG"
50681..51748
/genes="OSJNBa0095H06.8"
/join(50681..50683,51401..51748)
/genes="OSJNBa0095H06.8"

Query Match
Best Local Similarity 91.6%; Score 17.4; DB 8; Length 126447;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCTTCGCGAAAAAATCG 19
DB 45781 GCGTTCGCGAAAAAATCG 45763

RESULT 5
AB013090/c
LOCUS AB013090 2669 bp DNA linear BCT 19-JUN-1999
DEFINITION Clostridium histolyticum rrlB gene for 23S rRNA, partial sequence.
ACCESSION AB013090
VERSION AB013090.1 GI:3080567
KEYWORDS rrlB; 23S ribosomal RNA.
SOURCE Clostridium histolyticum
ORGANISM Clostridium; Clostridiales; Clostridiaceae;
REFERENCE 1 (sites)
AUTHORS Matsushita,O., Jung,C.M., Katayama,S., Minami,J., Takahashi,Y. and
Okabe,A.
TITLE Gene duplication and multiplicity of collagenases in Clostridium
histolyticum
JOURNAL J. Bacteriol. 181 (3), 923-933 (1999)
PUBMED 99121032
PUBMED 992257
REFERENCE 2 (bases 1 to 2669)
AUTHORS Matsushita,O.
DIRECT SUBMISSION
JOURNAL
TITLE
JOURNAL

FEATURES
source
1..2669
/organism="Clostridium histolyticum"
/mol_type="genomic DNA"
/strain="JCM 1403"
/db_xref="taxon:1498"
/dev_stage="vegetative cell"
1..2669
/genes="rrlB"
<1..>2669
/genes="rrlB"
/product="23S ribosomal RNA"

Query Match
Best Local Similarity 89.5%; Score 17; DB 1; Length 2669;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 CTTCGCGAAAAAATCG 19
DB 1241 CTTCGCGAAAAAATCG 1225

RESULT 6
AB013091/c
LOCUS AB013091 2669 bp DNA linear BCT 19-JUN-1999
DEFINITION Clostridium histolyticum rrlC gene for 23S rRNA, partial sequence.
ACCESSION AB013091
VERSION AB013091.1 GI:3080568
KEYWORDS rrlC; 23S ribosomal RNA.
SOURCE Clostridium histolyticum
ORGANISM Clostridium; Clostridiales; Clostridiaceae;
REFERENCE 1 (sites)
AUTHORS Matsushita,O., Jung,C.M., Katayama,S., Minami,J., Takahashi,Y. and
Okabe,A.
TITLE Gene duplication and multiplicity of collagenases in Clostridium
histolyticum
JOURNAL J. Bacteriol. 181 (3), 923-933 (1999)
PUBMED 99121032
PUBMED 992257
REFERENCE 2 (bases 1 to 2669)
AUTHORS Matsushita,O.
DIRECT SUBMISSION
JOURNAL
TITLE
JOURNAL

FEATURES
source
1..2669
/organism="Clostridium histolyticum"
/mol_type="genomic DNA"
/strain="JCM 1403"
/db_xref="taxon:1498"
/dev_stage="vegetative cell"
1..2669
/genes="rrlB"
<1..>2669
/genes="rrlB"
/product="23S ribosomal RNA"

Query Match
Best Local Similarity 100.0%; Score 17; DB 1; Length 2669;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 CTTCGCGAAAAAATCG 19
DB 1241 CTTCGCGAAAAAATCG 1225
```



```

REFERENCE
AUTHORS      Matsushita,O., Jung,C.M., Katayama,S., Minami,J., Takahashi,Y. and
              Okabe,A.
TITLE        Gene duplication and multiplicity of collagenases in Clostridium
              histolyticum
JOURNAL      J. Bacteriol. 181 (3), 923-933 (1999)
MEDLINE      99121032
PUBMED       9922257
REFERENCE    2 (bases 1 to 2669)
AUTHORS      Matsushita,O.
TITLE        Direct Submission
JOURNAL      Submitted (17-APR-1998) Osamu Matsushita, Kagawa Medical
              University, Department of Microbiology, 1750-1 Ikenobe, Miki-cho,
              Kagawa 761-0793, Japan (E-mail:osamu@kms.ac.jp,
              Tel:+81(87)891-2129, Fax:+81(87)891-2129)
FEATURES
source      Location/Qualifiers
              1..2669
              /organism="Clostridium histolyticum"
              /mol_type="genomic DNA"
              /strain="JCM 1403"
              /db_xref="taxon:1498"
              /dev_stage="vegetative cell"
              1..2669
              /gene="rrlC"
              <1..>2669
              /gene="rrlC"
              /product="23S ribosomal RNA"

gene
rrna        <1..>26751
              /gene="rrlA"
              /product="23S ribosomal RNA"

ORIGIN
Query Match      89.5%; Score 17; DB 1; Length 2669;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      3 CTTCGCGAAAAAATCG 19
      |||||
Db      1241 CTTCGCGAAAAAATCG 1225

RESULT 7
LOCUS      AB013089/c
DEFINITION Clostridium histolyticum rrlA gene for 23S ribosomal RNA
ACCESSION  AB013089
VERSION     AB013089.1 GI:3080566
KEYWORDS   rrlA; 23S ribosomal RNA.
SOURCE     Clostridium histolyticum
ORGANISM   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
              Clostridium.
REFERENCE
AUTHORS      Matsushita,O., Jung,C.M., Katayama,S., Minami,J., Takahashi,Y. and
              Okabe,A.
TITLE        Gene duplication and multiplicity of collagenases in Clostridium
              histolyticum
JOURNAL      J. Bacteriol. 181 (3), 923-933 (1999)
MEDLINE      99121032
PUBMED       9922257
REFERENCE    2 (bases 1 to 2751)
AUTHORS      Matsushita,O.
TITLE        Direct Submission
JOURNAL      Submitted (17-APR-1998) Osamu Matsushita, Kagawa Medical
              University, Department of Microbiology, 1750-1 Ikenobe, Miki-cho,
              Kagawa 761-0793, Japan (E-mail:osamu@kms.ac.jp,
              Tel:+81(87)891-2129, Fax:+81(87)891-2129)
FEATURES
source      Location/Qualifiers
              1..2751
              /organism="Clostridium histolyticum"
              /mol_type="genomic DNA"
              /strain="JCM 1403"
              /db_xref="taxon:1498"
              /dev_stage="vegetative cell"
              1..2751
              /gene="rrlA"

gene

```

```

rrna        <1..>2751
              /gene="rrlA"
              /product="23S ribosomal RNA"

ORIGIN
Query Match      89.5%; Score 17; DB 1; Length 2751;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      3 CTTCGCGAAAAAATCG 19
      |||||
Db      1241 CTTCGCGAAAAAATCG 1225

RESULT 8
LOCUS      AB079543/c
DEFINITION Schizosaccharomyces pombe ubr11 gene for ring finger protein,
              complete cds.
ACCESSION  AB079543
VERSION     AB079543.1 GI:18640082
KEYWORDS
SOURCE     Schizosaccharomyces pombe (fission yeast)
ORGANISM   Schizosaccharomyces pombe
              Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
              Schizosaccharomycetales; Schizosaccharomycetaceae;
              Schizosaccharomyces.
REFERENCE
AUTHORS      Kitamura,K., Katayama,S., Dhut,S., Sato,M., Watanabe,Y.,
              Yamamoto,M. and Toda,T.
TITLE        Phosphorylation of Mei2 and Stell by Pat1 kinase inhibits sexual
              differentiation via ubiquitin proteolysis and 14-3-3 protein in
              fission yeast
JOURNAL      Dev. Cell 1 (3), 389-399 (2001)
MEDLINE      21559218
PUBMED       11702950
REFERENCE    2 (bases 1 to 6159)
AUTHORS      Kitamura,K.
TITLE        Direct Submission
JOURNAL      Submitted (06-FEB-2002) Kenji Kitamura, Hiroshima University,
              Center for Gene Science, Kagamiyama 1-4-2, Higashi-Hiroshima,
              Hiroshima 739-8527, Japan (E-mail:kkita@hiroshima-u.ac.jp,
              Tel:81-824-24-6273, Fax:81-824-22-7184).
              Location/Qualifiers
              1..6159
              /organism="Schizosaccharomyces pombe"
              /mol_type="genomic DNA"
              /db_xref="taxon:4896"
              /note="see also Acc# Z97208"
              1..6159
              /gene="ubr11"
              1..6159
              /gene="ubr11"
              /note="similar to Ubr1"
              /codon_start=1
              /product="ring finger protein"
              /protein_id="BAB84668.1"
              /db_xref="GI:18640083"
              /translation="MQLHDPAPLSQSPASRLQKILLSEAKHAYDNESEKTHLLOE
              VFLSLNTNEDNWKYFLKEKPGATSDPRLSLQHSPECAQLQDKRSGSKVCHVF
              RAGEVIYRCKNGLDNTCLVCPFHATNHEGHTHSISQISGICDGPDEAWNV
              LNKCHNVDDDEQKPEVIPLEQHSIRTHILLDFILDVFSFVNLSKAQSTVG
              SILADEASRLSSAKYGVADRPCNVRVLMWDEVHTPDVAVGVSEALDSNTAFGL
              EVAQVDSIGHFAVATSASVHEAIRANAIKENLVNVRTARDFREDICGILLEWF
              DDLESHVGYFADYLIIVCDILKWSPGLEKPAKPRVFNPLPLEIVNDDSDDI
              YAREELDVIANLQDETGVTRIANLGDDEADMTDPTIAGDPHPLDNDNDLLD
              FETEREDIDDLTDEVMEETEEAEADYPVNRNTRQDDVDQISNERTSENETDESQN
              TENVDNPQTHTPVPIPTATQDVVTRPFPNSQLNLRLQIINARRPFAAVCVQS
              LREDYKGPHTIPSSYSFVSFSSILDFLLDFLKWRLGLSKLVVFPNRL
              LLFKRLMGIRFVHYRSLATFLPADRPDHSVMFLSVQFFTFPSLAEVVYDYLFL
              NINATTLISLQSNRPSTLFSDDIETPTIQLNRQVLRRTYNLPSLGLVLLQHPQV
              KLVVDVDTVYVHVIDLLRVFQGVIPQORAILSHVQWDFPHGKNILFVQKRVAMLSNT
              VSSCTQAPYERLFVAKIITSTHTPKLDAESLPLSCIPSSSLTNTFTQPLVPFSV"

FEATURES
source      Location/Qualifiers
              1..6159
              /organism="Schizosaccharomyces pombe"
              /mol_type="genomic DNA"
              /db_xref="taxon:4896"
              /note="see also Acc# Z97208"
              1..6159
              /gene="ubr11"
              1..6159
              /gene="ubr11"
              /note="similar to Ubr1"
              /codon_start=1
              /product="ring finger protein"
              /protein_id="BAB84668.1"
              /db_xref="GI:18640083"
              /translation="MQLHDPAPLSQSPASRLQKILLSEAKHAYDNESEKTHLLOE
              VFLSLNTNEDNWKYFLKEKPGATSDPRLSLQHSPECAQLQDKRSGSKVCHVF
              RAGEVIYRCKNGLDNTCLVCPFHATNHEGHTHSISQISGICDGPDEAWNV
              LNKCHNVDDDEQKPEVIPLEQHSIRTHILLDFILDVFSFVNLSKAQSTVG
              SILADEASRLSSAKYGVADRPCNVRVLMWDEVHTPDVAVGVSEALDSNTAFGL
              EVAQVDSIGHFAVATSASVHEAIRANAIKENLVNVRTARDFREDICGILLEWF
              DDLESHVGYFADYLIIVCDILKWSPGLEKPAKPRVFNPLPLEIVNDDSDDI
              YAREELDVIANLQDETGVTRIANLGDDEADMTDPTIAGDPHPLDNDNDLLD
              FETEREDIDDLTDEVMEETEEAEADYPVNRNTRQDDVDQISNERTSENETDESQN
              TENVDNPQTHTPVPIPTATQDVVTRPFPNSQLNLRLQIINARRPFAAVCVQS
              LREDYKGPHTIPSSYSFVSFSSILDFLLDFLKWRLGLSKLVVFPNRL
              LLFKRLMGIRFVHYRSLATFLPADRPDHSVMFLSVQFFTFPSLAEVVYDYLFL
              NINATTLISLQSNRPSTLFSDDIETPTIQLNRQVLRRTYNLPSLGLVLLQHPQV
              KLVVDVDTVYVHVIDLLRVFQGVIPQORAILSHVQWDFPHGKNILFVQKRVAMLSNT
              VSSCTQAPYERLFVAKIITSTHTPKLDAESLPLSCIPSSSLTNTFTQPLVPFSV"

gene
CDS

```

SRDPIFYHFLHMLNLSFYCYRDVASSHWKDTLLALLDHPURVCLVLAQIDCNLMI
 RNGRSILLTAFYRQNNIEVSXDKDILAIQITLMEFVNDENLVNAVQFETDMLYN
 LTYNEHNYETIRPAMLCQDELLIALITEREQILHVDIOIIRTRLAQCFGLA
 YSALLSTISNLSVSFDRIBREVSYKAPDGLHDFGVSLKDEYDVLQVYFHYN
 KNEREESDTILKRLAKNNVSAESIIEPKIRFLKGDHDIFFAAVNASTSLIFR
 AIEVALVOAESFGSSDIGNTILGDALQCLISMKIHEFKSNDFFCSRSCABRYPDSS
 IMREFGSAVCLAEILFAILKSPKVDVKNVAVLAGLQKNDPDSAYSNMLATPEL
 STVSSSDSNEIKETQEKLEALEKQKINQPRDQAQSLAQNTDFDGEQOTDEEV
 TTEPEBEVYKHIRGNCCLCEQNECDQAPYIGVIGLQSSLRKIDVHSEILIDEI
 YSVPPNDRSHRPFKGYDTVFNRSKDRLLSAPPNGNINRVFVSCGHLMLGC
 FKNYVARSYRNDVNTAGSEYKYSTAKFFMCLRSLSNVLPPQIPKWCINLD
 TLPFRSMGWLBEIGTMSSEFYQLVRSSLDTKDTRFSCFLRPWINSKIISAMLA
 RLKIALDQONNDRVSLDYRYCETTKLAWKVKGSTFTNVPDILNLSLAYTVS
 SLEVSORCSPKQSGATSVWFNLGELTTLKVLPLSDTVLKVCDQIISKDQOALLM
 ESOKLAVCKIFPYHSQKSLMNRGMSDHDQIOFLLSNFTDFDVKISSMLIFGQD
 NIIYVFLYSICKTIISMVKVADSSVVPVLLTINYSQQSKSYIILCKNVLLWCG
 SNNSIILDDENLLRLMSVYISPLFLRVALLYCMFDISLEFNESNDDSEL
 ERISKVPLPELQYSQMSDENQIILEIAGWCEHLAQNTGDSITISLEYPFGIEL
 VKLPHLEINIDMSQSVCCMCHKTPIALCMCGSVICFNARQNTVSSRLTQECN
 KHAATCTGSGVIGFPIFKACGILLDLSINTGTPMTPYLDIHGETDLQLRRCQCPFLN
 OKRYDFVVRQQLRQTVLQKARHMDTENGQNRMA"

ORIGIN

Query Match 89.5%; Score 17; DB 8; Length 6159;
 Best Local Similarity 100.0%; Pred. No. 3.8e+02;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 CTTCGCGAAAAAATCG 19

Db 973 CTTCGCGAAAAAATCG 957

RESULT 9

SPAC15A10/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

SPAC15A10 43702 bp DNA linear PLN 20-JUN-2003
 S.pombe chromosome I cosmid c15A10.
 Z97208

297208.1 GI:2239177

Schizosaccharomyces pombe (fission yeast)

Schizosaccharomyces pombe

Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

Schizosaccharomycetales; Schizosaccharomycetaceae;

Schizosaccharomycetes.

1 (bases 1 to 43702)

Wood, V., Gwilliam, R., Rajandream, M.A., Lyne, M., Lyne, R.,
 Stewart, A., Sgouros, J., Peat, N., Hayles, J., Baker, S., Basham, D.,
 Bowman, S., Brooks, K., Brown, D., Brown, S., Chillingworth, T.,
 Churcher, C., Collins, M., Connor, R., Cronin, A., Davis, P.,
 Feltwell, T., Fraser, A., Gentles, S., Goble, A., Hamlin, N., Harris, D.,
 Hidalgo, J., Hodgson, G., Holtroyd, S., Hornesby, T., Howarth, S., M.,
 Huckle, E.J., Hunt, S., Jagels, K., James, K., Jones, L., Jones, M.,
 Leather, S., McDonald, S., McLean, J., Mooney, P., Moule, S.,
 Mungall, K., Murphy, L., Niblett, D., Odell, C., Oliver, K., O'Neill, S.,
 Pearson, D., Quail, M.A., Rabinowitz, E., Rutherford, K., Rutter, S.,
 Saunders, D., Seeger, K., Sharp, S., Skelton, J., Simmonds, M.,
 Squires, R., Squires, S., Stevens, K., Taylor, K., Taylor, R.G.,
 Tivey, A., Walsh, S., Warren, T., Whitehead, S., Woodward, J.,
 Volkmer, G., Aert, R., Robben, J., Grymonprez, B., Weltjens, I.,
 Vanstreels, E., Rieger, M., Schafer, M., Muller-Auer, S., Gabel, C.,
 Fuchs, M., Dusterhoft, A., Fritzc, C., Holzer, E., Moestl, D.,
 Hilbert, R., Borzym, K., Langer, I., Beck, A., Lehrach, H.,
 Reinhardt, R., Pohl, T.M., Eger, P., Zimmermann, W., Wedler, H.,
 Wambutt, R., Purnelle, B., Goffeau, A., Cadieu, E., Dreano, S.,
 Gloux, S., Lelaure, V., Mortier, S., Galibert, F., Aves, S.J., Xiang, Z.,
 Hunt, C., Moore, K., Hurst, S.M., Lucas, M., Rochet, M., Gallard, C.,
 Tallada, V.A., Garzon, A., Thode, G., Daga, R., Cruzado, L.,
 Jimenez, J., Sanchez, M., del Rey, F., Benito, J., Dominguez, A.,
 Revuelta, J.L., Moreno, S., Armstrong, J., Forsburg, S.L., Cerutti, L.,
 Lowe, T., McCombie, W.R., Paulsen, I., Potashkin, J., Shpakovski, G.V.,
 Ussery, D., Barrell, B.G. and Nurse, P.
 The genome sequence of Schizosaccharomyces pombe
 Nature 415 (6874), 871-880 (2002)

MEDLINE

21848401

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

2 (bases 1 to 43702)
 Murphy, L., Harris, D., Wood, V., Barrell, B.G. and Rajandream, M.A.
 Direct Submission
 Submitted (27-JUN-1997) Schizosaccharomyces pombe chromosome I
 sequencing project, Sanger Institute, Hinxton Hall, Hinxton,
 Cambridge CB10 1HQ E-mail: pombe@sanger.ac.uk
 Notes:
 Details of S. pombe sequencing at the Sanger Institute are
 available on the World Wide Web.
 (URL, <http://www.genedb.org/genedb/pombe/index.jsp>) (URL,
http://www.sanger.ac.uk/Projects/S_pombe/)
 CDS are numbered using the following system eg SPAC5H10.01c. SP (S.
 pombe), A (chromosome 1), c5H10 (cosmid name), .01 (first CDS), c
 (complementary strand). However, clones may have been reorientated
 since the original submission, therefore the complementary strand
 notation may be invalid for strand inference. IMPORTANT: This
 sequence MAY NOT be the entire insert of the sequenced clone. It
 may be shorter because we only sequence overlapping sections once.

FEATURES

source

Location/Qualifiers

1..43702

/organism="Schizosaccharomyces pombe"

/mol_type="genomic DNA"

/strain="972h-"

/db_xref="taxon:4896"

/chromosome="I"

/map="ICEN"

/clone="cosmid c15A10"

1..11404

/gene="SPAC15A10.01"

/note="synonym: SPAC8C9.18"

<1..1404

/gene="SPAC15A10.01"

/codon_start=1

/protein_id="CAB10098.1"

/db_xref="GI:2239178"

/db_xref="GOA:O14286"

/db_xref="SWISS-PROT:O14286"

/translation="DRGTGKISFLSSMLHIPTILEIAMVSGILTYKGPSFAIA

ATTVALVFTVTRTWRVROANADSKASAAIESLYNEAVTFNNEVEMGR

VERHLSAVERANVAVSSLAFLNSGOALIFSTALTMYMGCGIVTSNLTGVGLVMI

NOVFOISIPLNLFLGVSREMGAFDMEOQLPSLKINIOVKRNPARDLVLKGGSIQ

FDNVHFSYNNRPLNCSFNIPAGKAVFVAGSGCGSKILLRLLFRFYDTSKILLI

DNRLDQITLNLKKAIGVVPDTPFLNFDILYNGYNKPNKSNDEIVEAKKAKIHD

LIETFPGYQTKVGERGIMSGGKQLAVSRLLKNPEILFFDEATSAIDTNTERAL

LRNFDILGKSHKTSVPFIAHRLTIKDCDIIFVLEKGRVVEQGSHEQLMAKNSVYTSM

WHQSEPFGRSNKSGDA"

1..1195

/gene="SPAC15A10.01"

/note="nominal overlap with cosmid SPAC8C9, EM:Z99168 S.

pombe chromosome 1"

727..1287

/gene="SPAC15A10.01"

/note="match to PF00005 ABC_tran, ABC transporter Score

219..90"

748..771

/gene="SPAC15A10.01"

/note="PS00017 ATP/GTP-binding site motif A (P-loop)"

1060..1104

/gene="SPAC15A10.01"

/note="PS00211 ABC transporters family signature"

1882..3234

/gene="SPAC15A10.02"

1882..3234

/gene="SPAC15A10.02"

/note="transcription initiation factor TFIID subunit

(predicted); TAF(II) complex (TBP-associated protein

complex); SAGA complex (predicted); involved in

establishment and/or maintenance of chromatin architecture

(predicted); similar to S. cerevisiae TAF61"

/codon_start=1

/protein_id="CAB10099.1"

```
/db_xref="GI:2239179"
/db_xref="GOA:O13722"
/translation="MNGCHSSPGTVPORPSAGPVGNOAQFSQORTNLTSLHTMTMYQ
OLAQNLNPGQVGLLOQTTRQLOQHAQSGQLPNOQLQLOASGALOGCTPEP
SNTRPQLNAQEGTLLVRRHQTAQNYLTKEKALGRINKELSTNERLDSAREA
LVKQSELVQAQTAASINGRISIOQLQRQASSANGNTGTFPVNASTDTRKST
ASTPQLOQANAPQANRPINPSTSSVPTFGVSAANVSNESTELATQOOSGLAN
NVSQPTSPYNSANHLKPKVSKGPIPFSPVPSRATLTGGYASIGLSTPGLSAPHY
ELDNGRLKRLHLLQIDSEKIEPEVESELLEIADEFVESVITFACRLAKHRK
SSTDVDVQLHLERNWIRLPGFASDDIVKSARKTGTPTPSYQKQNAIGTAKSLNKD
"
complement(4384..6942)
/gene="RAD54"
/note="RAD54"
/note="synonyms: rhp54, SPAC15A10.03c"
complement(4384..6942)
/gene="RAD54"
/note="DEAD/DEAH box helicase; SNF2 family; helicase
C-terminal domain; DNA dependent ATPase activity
(predicted); DNA supercoiling activity (predicted);
involved in DNA repair; deletion mutant sensitive to UV
(PMID 8834792); deletion mutant sensitive to ionizing
radiation (PMID 8834792); deletion mutant results in
elongated cells (frequent) (PMID 8834792); deletion mutant
results in aberrant nuclei (occasional) (PMID 8834792);
deletion mutant results in a high level of chromosome loss
(PMID 8834792); involved in meiotic recombination (PMID
8834792); involved in DNA repair; localization nucleus
(predicted); involved in chromatin remodeling (predicted);
involved in heteroduplex formation (predicted); similar to
S. cerevisiae RAD54"
/codon_start=1
/protein_id="CAB10100.1"
/db_xref="GI:2239180"
/db_xref="GOA:P41410"
/db_xref="SWISS-PROT:P41410"
/translation="MIQOFTAKPRISTSSKLTVLKNKNVPGKLFKKFKPSLVI
SEKKELELRKPRVNSYSGVDKSDAYSVNSGLATIKKANRLINLHRRRPS
TVIKQFQVPIKIHEDISKLCARPPPTLGMKRVDFPRPLYDPADFAFLVLYDP
TDADEILPDIKEVLAERKDELLKNGKKEISDSEPSDSDSCVSTVASCSTE
QSLTTSKRRRPNKSLDLGQKEKPPPPVAVVDPKLIARLPHOIEGVKPLY
KQVTRIDRCANGICMDEMGLKTLQALITLWLLQSKQPAQKPTIEKAIITCPSSL
VKNANELVWLKDAITPFLIDGSSKQELIMALQWASVHGQVTRPVLIASVETL
RSYVHLNNAEIGMLDCDEHRLKNSLITFALDNLNVRVILSTPQNDLSYF
SLINPANGPLIGSGQFKNYEIFILKGRDAGTEKDKENGDAKLAELAKIVARFIR
RTNDILSLPVKYEYVFNLSFQLSLRYHFTSPINKILGTSQPLKAIGLIK
KICNPDLNLTDELEGCEALFPFGFIPRELGRDRNIDSSLSGMLVLERMLYQIK
ETDDKIVLISNTVTLDFEOLCRARGYKALRLDGTWNKRLVTFDNDPEKDAFY
FLSSKAGCGINLIGANRLILFDPDNPADQALARVWDGKKDFVRYFIATGT
LEETKFORQSHKSLSSCVDEADQVHERHESLNLRLQLFOLNHTVCTHETVCKRC
RDGQFIAPAMLVGDRSTWNHFTNFDLIEDHLLKREAGKQGVSVFYQKSH"
complement(4768..5022)
/gene="RAD54"
/note="RAD54"
/note="Match to PF00271 helicase_C, Helicases conserved
C-terminal domain Score 106.44"
complement(5218..6144)
/gene="RAD54"
/note="Match to PF00176 SNF2_N, SNF2 and others N-terminal
domain Score 500.99"
7449..7568
/note="mRNA from AU013639"
complement(7519..8898)
/gene="SPAC15A10.04c"
/note="synonym: zpr1"
complement(7519..8898)
/gene="SPAC15A10.04c"
/note="zinc finger protein; possibly binds to translation
elongation factor (PMID 9852145); similar to S. cerevisiae
ZPR1"
/codon_start=1
/protein_id="CAB10101.1"
/db_xref="GI:2239181"
/db_xref="GOA:O13724"
/db_xref="SWISS-PROT:O13724"
/translation="MAEKEKEELFTSINAAQNVSTAEADREGNVQVSESLCMCGKN
GTTKLLLTIVFYREVVLMSPECFGFAQVQHAETIOPEGSKIIFHVEDKEDLNR
TVVKSQAIVSIPBIOLEIPGRIGLATTIEILSNVVDLDSKEQSKESAPLYDQI
NAFKVNSLSSGSPVPTITVDDITGNSWIEMKPGRDGRWSQVSKYRTEPQNTKGL
VTPQPEDVKTQTNASNTLKHDATAVEVDENRVTHTATCPSCSHOCDTHMKLLDIP
HFKEVIMSTVCDRCGVRSEVKTGGIPEPKRKITLKWDAEQLDSRDLKSETASLK
IPEGLDLFPPTLGLGRFTTIEGLAQVITDIEYGRVFSQETDSMTPEQVANWQQLCNL
TAAREGATQFTTILDDPLSQSYLYAPDDPNTTIEEYERSQVNEELGNDKTE
NYEKDGGRK"
complement(join(9557..9786,9871..10286,10327..10397,
10507..10518))
/gene="SPAC15A10.05c"
complement(join(9557..9786,9871..10286,10327..10397,
10507..10518))

Query Match 89.5%; Score 17; DB 8; Length 43702;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 CTTCGCGAARAAATCG 19
Db 27564 CTTCGCGAARAAATCG 27548

RESULT 10
AC079046 132930 bp DNA linear HTG 17-AUG-2000
LOCUS Homo sapiens chromosome 3 clone RP11-744J3 map 3, LOW-PASS SEQUENCE
DEFINITION SAMPLING.
AC079046.1 GI:9837971
VERSION HTG; HTGS; PHASE0.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 132930)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 3, clone RP11-744J3
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 132930)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bada,F., Boguslavsky,L.,
Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G.,
Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mienga,V.,
Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
O'Donnell,P., O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Travers,M., Trigglio,J., Vassiliev,H., Visl,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (17-AUG-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green,P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
```

----- Project Information
Center project name: L10218
Center clone name: 744_J_3

* NOTE: This record contains 158 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1 743: contig of 743 bp in length
* 744 843: gap of 100 bp
* 844 1589: contig of 746 bp in length
* 1590 1689: gap of 100 bp
* 1690 2437: contig of 748 bp in length
* 2438 2537: gap of 100 bp
* 2538 3287: contig of 750 bp in length
* 3288 3387: gap of 100 bp
* 3388 4133: contig of 746 bp in length
* 4134 4233: gap of 100 bp
* 4234 4963: contig of 730 bp in length
* 4964 5063: gap of 100 bp
* 5064 5785: contig of 722 bp in length
* 5786 5885: gap of 100 bp
* 5886 6615: contig of 730 bp in length
* 6616 6715: gap of 100 bp
* 6716 7462: contig of 747 bp in length
* 7463 7562: gap of 100 bp
* 7563 8310: contig of 748 bp in length
* 8311 8410: gap of 100 bp
* 8411 9144: contig of 734 bp in length
* 9145 9244: gap of 100 bp
* 9245 9985: contig of 741 bp in length
* 9986 10085: gap of 100 bp
* 10086 10831: contig of 746 bp in length
* 10832 10931: gap of 100 bp
* 10932 11677: contig of 746 bp in length
* 11678 11777: gap of 100 bp
* 11778 12516: contig of 739 bp in length
* 12517 12616: gap of 100 bp
* 12617 13329: contig of 713 bp in length
* 13330 13429: gap of 100 bp
* 13430 14149: contig of 720 bp in length
* 14150 14249: gap of 100 bp
* 14250 14980: contig of 731 bp in length
* 14981 15080: gap of 100 bp
* 15081 15809: contig of 729 bp in length
* 15810 15909: gap of 100 bp
* 15910 16655: contig of 746 bp in length
* 16656 16755: gap of 100 bp
* 16756 17520: contig of 785 bp in length
* 17521 17621: gap of 100 bp
* 17622 18377: contig of 757 bp in length
* 18378 18477: gap of 100 bp
* 18478 19229: contig of 752 bp in length
* 19230 19329: gap of 100 bp
* 19330 20059: contig of 730 bp in length
* 20060 20159: gap of 100 bp
* 20160 20887: contig of 728 bp in length
* 20888 20987: gap of 100 bp
* 20988 21731: contig of 744 bp in length
* 21732 21831: gap of 100 bp
* 21832 22568: contig of 737 bp in length
* 22569 22668: gap of 100 bp
* 22669 23408: contig of 740 bp in length
* 23409 23508: gap of 100 bp
* 23509 24262: contig of 754 bp in length
* 24263 24362: gap of 100 bp

24363 25106: contig of 744 bp in length
25107 25206: gap of 100 bp
25207 25955: contig of 749 bp in length
25956 26055: gap of 100 bp
26056 26802: contig of 747 bp in length
26803 26902: gap of 100 bp
26903 27657: contig of 755 bp in length
27658 27758: gap of 100 bp
27759 28473: contig of 716 bp in length
28474 28573: gap of 100 bp
28574 29295: contig of 722 bp in length
29296 29395: gap of 100 bp
29396 30114: contig of 719 bp in length
30115 30214: gap of 100 bp
30215 30963: contig of 748 bp in length
30964 31062: gap of 100 bp
31063 31814: contig of 752 bp in length
31815 31914: gap of 100 bp
31915 32659: contig of 745 bp in length
32660 32759: gap of 100 bp
32760 33506: contig of 747 bp in length
33507 33606: gap of 100 bp
33607 34373: contig of 767 bp in length
34374 34473: gap of 100 bp
34474 35240: contig of 767 bp in length
35241 35340: gap of 100 bp
35341 36100: contig of 760 bp in length
36101 36200: gap of 100 bp
36201 36943: contig of 743 bp in length
36944 37043: gap of 100 bp
37044 37765: contig of 722 bp in length
37766 37865: gap of 100 bp
37866 38593: contig of 728 bp in length
38594 38693: gap of 100 bp
38694 39418: contig of 725 bp in length
39419 40250: contig of 732 bp in length
40251 40350: gap of 100 bp
40351 41062: contig of 712 bp in length
41063 41162: gap of 100 bp
41163 41909: contig of 747 bp in length
41910 42009: gap of 100 bp
42010 42765: contig of 756 bp in length
42766 42865: gap of 100 bp
42866 43606: contig of 741 bp in length
43607 43706: gap of 100 bp
43707 44450: contig of 744 bp in length
44451 44550: gap of 100 bp
44551 45266: contig of 715 bp in length
45266 45365: gap of 100 bp
45366 46078: contig of 713 bp in length
46079 46178: gap of 100 bp
46179 46923: contig of 745 bp in length
46924 47023: gap of 100 bp
47024 47761: contig of 738 bp in length
47762 47861: gap of 100 bp
47862 48593: contig of 732 bp in length
48594 48693: gap of 100 bp
48694 49447: contig of 754 bp in length
49448 49547: gap of 100 bp
49548 50304: contig of 757 bp in length
50305 50404: gap of 100 bp
50405 51148: contig of 744 bp in length
51149 51248: gap of 100 bp
51249 51997: contig of 749 bp in length
51998 52097: gap of 100 bp
52098 52853: contig of 756 bp in length
52854 52953: gap of 100 bp
52954 53669: contig of 716 bp in length
53670 53769: gap of 100 bp
53770 54488: contig of 719 bp in length
54489 54588: gap of 100 bp
54589 55316: contig of 728 bp in length

```

* 55317 55416: gap of 100 bp
* 55417 56182: contig of 766 bp in length
* 56183 56282: gap of 100 bp
* 56283 57020: contig of 738 bp in length
* 57021 57120: gap of 100 bp
* 57121 57881: contig of 761 bp in length
* 57882 57981: gap of 100 bp
* 57982 58738: contig of 757 bp in length

Query Match      89.5%; Score 17; DB 2; Length 132930;
Best Local Similarity 100.0%; Pred. No. 4.9e+02; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 0;

QY 1 GCCTTCGCGAAAAAAT 17
DB 90954 GCCTTCGCGAAAAAAT 90970

RESULT 11
LOCUS CNS08CA3 157690 bp DNA linear PLN 21-NOV-2003
DEFINITION Oryza sativa chromosome 12, . BAC OJ1396_C02 of library Monsanto
from chromosome 12 of cultivar Nipponbare of ssp. japonica of Oryza
sativa (rice), complete sequence.
ACCESSION AL772425
VERSION AL772425.7 GI:37718622
KEYWORDS HTG.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
1 (bases 1 to 157690)
Choisne,N., Orjeda,G., Catolico,L., Demange,N., Wincker,P.,
Segurens,B., Pelletier,E., Scarpelli,C., Salanoubat,M.,
Weissenbach,J. and Quetier,F.
Oryza sativa chromosome 12 sequencing
Unpublished
Oryza sativa chromosome 12 sequencing
Genoscope
2 (bases 1 to 157690)
Direct Submission
Submitted (20-NOV-2003) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
On Oct 17, 2003 this sequence version replaced gi:34556135.
Center: Genoscope / Centre National de Sequencage
Center code: GS
Web site: http://www.genoscope.cns.fr/
Contact: SeqRef@genoscope.cns.fr
-----
The following sequence is oriented from the T7 to the SP6 end. The
nucleotide sequence of this BAC clone was generated by combining
Monsanto, Syngenta and Genoscope sequencing data.
Upstream BAC (overlapping the T7 end) : OSJNba0018E22 (AC=AL845346)
Downstream BAC (overlapping the SP6 end) : OSJNba0036A19
(AC=AL844878) ----- Finishing boundaries
FINISHED SEGMENT STARTS AT BASE 1
FINISHED SEGMENT ENDS AT BASE 157690
-----
FEATURES
source
1. 157690
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/sub_species="japonica"
/db_xref="taxon:39947"
/chromosome="12"
/clone="OJ1396_C02"
/clone_lib="Monsanto"

Query Match      89.5%; Score 17; DB 8; Length 157690;
Best Local Similarity 100.0%; Pred. No. 4.9e+02;

```

```

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCTTCGCGAAAAAAT 17
DB 133945 GCCTTCGCGAAAAAAT 133961

RESULT 12
LOCUS AX887299/c 386 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 3162 from Patent EP1033401.
ACCESSION AX887299
VERSION AX887299.1 GI:40045321
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Dumas Milne Edwards,J.B., Duclert,A. and Giordano,J.Y.
Expressed sequence tags and encoded human proteins
Patent: EP 1033401-A 3162 06-SEP-2000;
Genset (FR)
FEATURES
source
1..386
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
139..>384
/note="unnamed protein product"
/codon_start=1
/protein_id="CAP01437.1"
/db_xref="GI:40045322"
/translation="MNPFHSCWNTSABLKNKPEAYOTASVVDTVILPMSIGII
CSTGLVGNILIVFTIIRSKKTVDPDIYICNLAVADLVH"

ORIGIN
Query Match      86.3%; Score 16.4; DB 6; Length 386;
Best Local Similarity 94.4%; Pred. No. 6.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCTTCGCGAAAAAATC 18
DB 47 GCCTTCGCGAATAAATC 30

RESULT 13
LOCUS BD026909/c 386 bp DNA linear PAT 27-AUG-2002
DEFINITION Sequence tag and encoded human protein.
ACCESSION BD026909
VERSION BD026909.1 GI:22568651
KEYWORDS JP 2001269182-A/3155.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 386)
Edwards,J.B.D.M., Duclair,E. and Jordan,J.Y.
Sequence tag and encoded human protein
Patent: JP 2001269182-A 3155 02-OCT-2001;
GENSET
COMMENT
OS Homo sapiens (human)
PN JP 2001269182-A/3155
PD 02-OCT-2001
PF 24-FEB-2000 JP 2000118773
PR 26-FEB-1999 US 60/122487
PI JEAN BAPTISTE DUMAS MILNE EDWARDS,EIMERIC DUCLAIR,JEAN YVES
PI JORDAN
PI C12N15/09,C07K14/435,C07K16/18,C12N1/15,C12N1/19,C12N1/21, PC
C12N5/10,
PC C12P21/02,C12P21/08,C12Q1/68//G06F17/30,C12N15/00,C12N5/00, PC
G06F15/40

```

```

CC      Key      Location/Qualifiers
FH      CDS
FT      CDS      139..384.
FEATURES             Location/Qualifiers
     source          1..386
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
ORIGIN
Query Match      86.3%; Score 16.4; DB 6; Length 386;
Best Local Similarity 94.4%; Pred. No. 6.e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY      1 GCCTTCGCGAATAAATC 18
Db      47 GCCTTCGCGAATAAATC 30

RESULT 14
AE008752
LOCUS      Salmonella typhimurium LT2, section 56 of 220 of the complete genome. BCT 23-APR-2003
DEFINITION      Salmonella typhimurium LT2
ACCESSION      AE008752 AE006468
VERSION        AE008752.1 GI:16419709
KEYWORDS
SOURCE
ORGANISM      Salmonella typhimurium LT2
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
REFERENCE      1 (bases 1 to 20097)
AUTHORS      McClelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W.,
Latreille, P., Courtney, L., Porwollik, S., Ali, J., Dante, M., Du, F.,
Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A.,
Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Florea, L., Miller, W.,
Stonking, T., Nhan, M., Waterston, R., and Wilson, R.K.
Complete genome sequence of Salmonella enterica serovar Typhimurium.
LT2
JOURNAL      Nature 413 (6858), 852-856 (2001)
MEDLINE      21534948
PubMed      11677609
REFERENCE      2 (bases 1 to 20097)
AUTHORS
The Salmonella typhimurium Genome Sequencing Project
Direct Submission
Submitted (29-MAR-2001) Genome Sequencing Center, Department of
Genetics, Washington University School of Medicine, 4444 Forest
Park Boulevard, St. Louis, MO 63108, USA
COMMENT      Supported by NIH grant 5U 01 AI43283
Coding sequences below are predicted from manually evaluated
computer analysis, using similarity information and the programs;
Glimmer; http://www.tigr.org/softlab/glimmer/glimmer.html and
GeneMark; http://opal.biology.gatech.edu/GeneMark/
EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/,
and Pedro Romero and Peter Karp at EcoCyc;
http://ecocyc.farnsworthsystems.com/ecocyc/
The analyses of ribosome binding sites and promoter binding sites
were kindly provided by Heladia Salgado, Julio Collado-Vides and
ReguonDB;
http://kinich.cifn.unam.mx:8850/db/regulondb\_intro.frameset
This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate
chemistries or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one m13 subclone.
Location/Qualifiers
source
1..20097
/organism="Salmonella typhimurium LT2"
/mol_type="genomic DNA"
/strain="LT2; SGSC 1412; ATCC 700720"
/db_xref="ATCC:700720"
/db_xref="taxon:99287"
/ncfe="LT2"
71..1024
/gene="fabH"
/ncfe="synonym: STM1193"
71..1024
/gene="fabH"
/EC number="2.3.1.41"
/ncfe="3-oxoacyl-[acyl-carrier-protein] synthase III.
(SW: FABH_SALTY)"
/codon_start=1
/transl_table=11
/product="3-oxoacyl-[acyl-carrier-protein] synthase III;
acetylCoA ACP transacylase"
/protein_id="AAL20122.1"
/db_xref="GI:16419710"
/transl_table="MYTKIIGTGYLPEQVETNADLEKMYETSDWIVTGTGPKRHI
AAPNETVATMGFTAAAPAIEMAGIDKQIGLIVVATTSATHAPSAACIQSMKIGK
CPAPDVAACAGFTYALSIADQYKSGAVKHALVVGSDVLARTCDPDGRGTIIIFGG
AGAAVSASAEPEGIIISHLHADRYGELLTPNADRVNPDNPYLTWAGNEVFKAVT
ELAHIVDETLAANNDRSELNWLPHQANLRIISATAKLGMSMDNVVVTIDRHGNTS
AASVPCALDEAVDRGRIKAGQLVLEAFGGGTGWSALIRF"
1028..1969
/gene="fabD"
/ncfe="synonym: STM1194"
1028..1033
/gene="fabD"
/ncfe="fabD"
1040..1969
/ncfe="fabD"
/EC number="2.3.1.39"
/ncfe="malonyl coA-acyl carrier protein transacylase.
(SW: FABD_SALTY)"
/codon_start=1
/transl_table=11
/product="malonyl-CoA-[acyl-carrier-protein] transacylase"
/protein_id="AAL20123.1"
/db_xref="GI:16419711"
/transl_table="MTQFAFVFPQGSQSVGMLAEMAANYPIVEETFAEASALGYDL
WALTQGFABEELNKTQTPALLTASVALMRVWQGGKQPMALMAGSLGEISALVCA
GVINFADAVELVEMRGKMQEAVPEGTGMSAIIGLDASIAKACEESQVVPVN
FNSPGQVVIAGHKEAVERAGAKAKALEPLVSPSHCALMFPADKLAVELAK
ITFSAPTPVNVNVDKCECTDAAIIRDALVRLYNPQWTKSVEFIAAQGVHLYEYG
PGKVLTLTKRIVDTLTALNEPALSAALTQ"
1972..2716
/gene="fabG"
/ncfe="synonym: STM1195"
1972..1977
/gene="fabG"
/ncfe="fabG"
/ncfe="putative RBS for fabG; RegulonDB:STMS1H001455"
1982..2716
/gene="fabG"
/EC number="1.1.1.100"
/ncfe="3-oxoacyl-[acyl-carrier protein] reductase.
(SW: FABG_SALTY)"
/codon_start=1
/transl_table=11
/product="3-oxoacyl-[acyl-carrier-protein] reductase"
/protein_id="AAL20124.1"
/db_xref="GI:16419712"
/transl_table="MSFEGKIALVTGASRGIGRAIAETLVARGAKVIGTATSENGAKN
ISDYLKANGKLMNVTDPAISVLENIKAEFGEVDILVNNAGITRDNLMMKODE
WNIIEITNLSVFRSLKAVNRMMKRCRGIITIGSVVGTMGAGNAAKAGLIG
FSSKLAREVASRGITVNVVAPGFIETDMTRALSDQDRAGIQAQVPAGRLGGAETIASA
VAFLASDEASYITGETLHVNGGMVMV"
2856..3108
/gene="acpP"

```

```
RBS      /note="synonym: STM1196"
2856. .2861
/ gene="acpP"
/ note="putative RBS for acpP; RegulonDB:STMS1H001456"
CDS      2872. .3108
/ gene="acpP"
/ note="similar to E. coli acyl carrier protein
(AAC74178.1); Blastp hit to AAC74178.1 (78 aa), 100%
identity in aa 1 - 78"
/ codon_start=1
/ transl_table=11
/ product="acyl carrier protein"
/ protein_id="AAL20125.1"
/ db_xref="GI:16419713"
/ translation="MSTIERVKIIIGELGVQKEVTVNNAFVEDLGADSLDTVELV
MALEEFDTIPDEAEKITTQAAIDYINGHOA"
gene      3181. .4435
/ gene="fabF"
/ note="synonym: STM1197"
RBS      3181. .3186
/ gene="fabF"
CDS      / note="putative RBS for fabF; RegulonDB:STMS1H001457"
3194. .4435
/ gene="fabF"
/ EC number="2.3.1.41"
/ note="similar to E. coli 3-oxoacyl-(acyl-carrier-protein)
synthase II (AAC74179.1); Blastp hit to AAC74179.1 (413
aa), 94% identity in aa 1 - 413"
/ codon_start=1
/ transl_table=11
/ product="3-oxoacyl-(acyl-carrier-protein) synthase II"
/ protein_id="AAL20126.1"
/ db_xref="GI:16419714"
/ translations="MSKRRVVVTLGLMLSPVGNVTSTWKALLAGOSGLSLIDHEDTS
AYATKAGLVKDFNCDDIISREKQKMDAIFQYGVAGVQAMDSGLEVTENASRIG
AATGSGIGGLIEHSHSLYKGPRIKISPPFVPSTIVNVAHLITMIGLGPISII
ATACISVHNGHGAARLIAYGDADAMVAGAEKAKIPLGVGFGAALSTNDNPPQA
ASRPDKERDGFVLGAGMLVLEBEYHAKARAKIYAEIVGFGSSDAYHTMTPPEN
GAGAALAMVALRDAIAIEPAQIGYVNAHGTSTPAGKAEIQAQVKGSDAARVMVSS
TKGNTGHLGAAGAVSIFSIILALRDQAIPPTINLNDPDEGCDLDFVPEARQVSDLE
YALCNSFGFGTNGSLIFKKI"
4548. .5368
/ gene="pabC"
/ note="synonym: STM1198"
4548. .4553
/ gene="pabC"
CDS      / note="putative RBS for pabC; RegulonDB:STMS1H001458"
4559. .5368
/ gene="pabC"
/ EC number="4.-.-"
/ note="similar to E. coli 4-amino-4-deoxychorismate lyase
(AAC74180.1); Blastp hit to AAC74180.1 (269 aa), 69%
identity in aa 1 - 269"
/ codon_start=1
/ transl_table=11
/ product="4-amino-4-deoxychorismate lyase"
/ protein_id="AAL20127.1"
/ db_xref="GI:16419715"
/ translation="MFLINGHAQDLAVSDRATQFGDGSFTTARIVDGNICHLAEHLQ
RLQVCKELRIAFSHWSTLKEQMTLATGDSGLVKIILSRGSGGYSAMQCQATR
ILSVSPAYISQWRKGITLTLSPIPLGNPYLAGLKHNLLEQLIRSHLEQTDAD
EALVLDSEGVWTECCANLFWRTGDIYFTPLDQAGVNGIMRQFCILRQAQSPFOVLE
VQAREAVROARDIIICNALMPIIPIRAVHGTSTYSRNTLFOFLAPPCHEPN"
5371. .6393
/ gene="yceG"
/ note="synonym: STM1199"
5371. .6393
/ gene="yceG"
/ note="similar to E. coli putative thymidylate kinase
(AAC74181.1); Blastp hit to AAC74181.1 (340 aa), 85%
identity in aa 1 - 340"
/ codon_start=1
/ transl_table=11
```

```
/product="putative periplasmic solute-binding protein"
/protein_id="AAL20128.1"
Query Match      86.3%; Score 16.4; DB 1; Length 20097;
Best Local Similarity 94.4%; Pred. No. 9e+02; 1; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY      2 CTTTCGCGAAAAAATCG 19
      |||||
DB      16769 CTTTCGCGAAAAAATCG 16786
      |||||

RESULT 15
AC145319/c
LOCUS      42703 bp DNA linear HTG 24-JUN-2003
DEFINITION      Oryza sativa (japonica cultivar-group) chromosome 11 clone
OSUNBA0030J08, *** SEQUENCING IN PROGRESS ***, 6 unordered pieces.
AC145319
AC145319.1 GI:32171257
VERSION      HTG; HTGS_PHASE1; HTGS_ACTIVESTIN
KEYWORDS      Oryza sativa (japonica cultivar-group)
SOURCE      Oryza sativa (japonica cultivar-group)
ORGANISM      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE      1 (bases 1 to 42703)
AUTHORS      Buell,R., Hsiao,J., Zismann,V., Moffat,K.M., Hill,J.,
Gansberger,K., Burgess,S., Jarrahi,B., Shvartsbeyn,M., Brenner,M.,
Ciecko,A., Pai,G., Vanaken,S., Hansen,C., Utterbach,T.,
Feidblyum,T., Khalak,H.G., Yuan,Q., Quackenbush,J., White,O.,
Salzberg,S. and Fraser,C.
TITLE      Oryza sativa ssp. japonica cv. Nipponbare OSUNBA0030J08 BAC genomic
sequence
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 42703)
AUTHORS      Buell,R.
DIRECT SUBMISSION
JOURNAL      Submitted (24-JUN-2003) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
COMMENT      * NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 7657: contig of 7657 bp in length
* 7658 7757: gap of unknown length
* 7758 9407: contig of 1650 bp in length
* 9408 9507: gap of unknown length
* 9508 17088: contig of 7581 bp in length
* 17089 17188: gap of unknown length
* 17189 26022: contig of 8834 bp in length
* 26023 26122: gap of unknown length
* 26123 30604: contig of 4482 bp in length
* 30605 30704: gap of unknown length
* 30705 42703: contig of 11999 bp in length.
FEATURES
source
1. .42703
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/sub_species="japonica"
/db_xref="taxon:39947"
/chromosome="11"
/clone="OSUNBA0030J08"

ORIGIN
Query Match      86.3%; Score 16.4; DB 2; Length 42703;
Best Local Similarity 94.4%; Pred. No. 9.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

OY 1 GCCTTCGCGAAAAAATC 18
DB 585 GCCTTCGCGAAAAAATC 568

Search completed: April 24, 2004, 19:25:34
Job time : 566.494 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2004, 17:01:15 ; Search time 160.679 Seconds
(without alignments)
502.342 Million cell updates/sec

Title: US-10-084-555A-118
Perfect score: 19
Sequence: 1 gcttcgcgaataatcg 19

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq 29Jan04:*
1: Geneseqn1990s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002s:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	16.4	86.3	386	3 AAC03164	AAC03164 Human sec
C 2	16.4	86.3	703	6 ABQ52333	ABQ52333 Oligonuc1
C 3	16.4	86.3	703	6 ABQ52332	ABQ52332 Oligonuc1
C 4	16.4	86.3	720	6 ABQ37429	ABQ37429 Oligonuc1
C 5	16.4	86.3	720	6 ABQ37428	ABQ37428 Oligonuc1
C 6	16.4	84.2	4487	1 AAB60943	AAB60943 Sequence
C 7	15.8	83.2	134	7 AAB65329	AAB65329 Human gen
C 8	15.8	83.2	327	2 AAB28603	AAB28603 AcMNPV gp
C 9	15.8	83.2	369	7 AAB252684	AAB252684 Aspergill
C 10	15.8	83.2	522	6 ABQ47425	ABQ47425 Oligonuc1
C 11	15.8	83.2	522	6 ABQ47424	ABQ47424 Oligonuc1
C 12	15.8	83.2	524	6 ABQ47198	ABQ47198 Oligonuc1
C 13	15.8	83.2	524	6 ABQ47199	ABQ47199 Oligonuc1
C 14	15.8	83.2	600	6 ABQ27793	ABQ27793 Oligonuc1
C 15	15.8	83.2	600	6 ABQ27792	ABQ27792 Oligonuc1
C 16	15.8	83.2	827	6 ABQ32512	ABQ32512 Oligonuc1
C 17	15.8	83.2	827	6 ABQ32513	ABQ32513 Oligonuc1
C 18	15.8	83.2	827	6 ABQ35126	ABQ35126 Oligonuc1
C 19	15.8	83.2	827	6 ABQ35127	ABQ35127 Oligonuc1
C 20	15.8	83.2	850	6 ABQ49804	ABQ49804 Oligonuc1
C 21	15.8	83.2	850	6 ABQ49805	ABQ49805 Oligonuc1
C 22	15.8	83.2	992	6 ABQ45389	ABQ45389 Oligonuc1
C 23	15.8	83.2	992	6 ABQ45388	ABQ45388 Oligonuc1

C 24	15.8	83.2	996	6 ABQ42152	ABQ42152 Oligonuc1
C 25	15.8	83.2	996	6 ABQ42153	ABQ42153 Oligonuc1
C 26	15.8	83.2	1043	6 ABQ48868	ABQ48868 Oligonuc1
C 27	15.8	83.2	1043	6 ABQ48869	ABQ48869 Oligonuc1
C 28	15.8	83.2	2501	7 ABZ10073	ABZ10073 Haematopo
C 29	15.8	83.2	3234	6 ABK74877	ABK74877 Bacillus
C 30	15.8	83.2	4501	7 ADA20436	ADA20436 Prostata
C 31	15.8	83.2	4501	7 ADA84243	ADA84243 Human ren
C 32	15.8	83.2	8245	4 AAS46447	AAS46447 Tumour su
C 33	15.8	83.2	39824	9 ADC00831	ADC00831 Enterohae
C 34	15.8	83.2	48908	8 ACD19124	ACD19124 E. coli 0
C 35	15.4	81.1	244	3 AAC55136	AAC55136 Arabidops
C 36	15.4	81.1	535	6 ABQ15410	ABQ15410 Oligonuc1
C 37	15.4	81.1	535	6 ABQ15411	ABQ15411 Oligonuc1
C 38	15.4	81.1	609	6 ABQ28751	ABQ28751 Oligonuc1
C 39	15.4	81.1	609	6 ABQ28750	ABQ28750 Oligonuc1
C 40	15.4	81.1	621	6 ABQ38516	ABQ38516 Oligonuc1
C 41	15.4	81.1	621	6 ABQ38517	ABQ38517 Oligonuc1
C 42	15.4	81.1	904	6 ABQ20116	ABQ20116 Oligonuc1
C 43	15.4	81.1	904	6 ABQ20117	ABQ20117 Oligonuc1
C 44	15.4	81.1	911	6 ABQ52534	ABQ52534 Oligonuc1
C 45	15.4	81.1	911	6 ABQ52535	ABQ52535 Oligonuc1

ALIGNMENTS

RESULT 1
AAC03164/c
ID AAC03164 standard; cDNA; 386 BP.
XX
AC AAC03164;
XX
DT 06-OCT-2000 (first entry)
DE Human secreted protein 5' EST, SEQ ID NO: 3162.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
FN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-00200610.
XX
PR 26-FEB-1999; 99US-0122487P.
XX (GEST) GENSET.
Dumas Milne Edwards J, Duclert A, Giordano J;
WPI; 2000-500381/45.
P-PSDB; AAG03158.
New nucleic acid that is a 5' expressed sequence tag (5' EST) for obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for diagnostic, forensic, gene therapy and chromosome mapping procedures.
Claim 1; SEQ ID NO 3162; 71pp + Sequence Listing; English.
The present sequence is one of a large number of 5' ESTs derived from mRNAs encoding secreted proteins. An ORF has been identified within the sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs derived from 30 different tissues. EST sequences usually correspond CC mainly to the 3' untranslated region (UTR) of the mRNA because they are often obtained from oligo-dT primed cDNA libraries. Such ESTs are not well suited for isolating cDNA sequences derived from the 5' ends of mRNAs and even in those cases where longer cDNA sequences have been obtained, the full 5' UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be used to obtain full length

Query Match 86.3%; Score 16.4; DB 6; Length 703;
Best Local Similarity 94.4%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCTTCGCGAAAAAATCG 19
DB 94 CCTTCGCGAAAAAATCG 77

RESULT 4
ABQ37429
ID ABQ37429 standard; DNA; 720 BP.
XX
AC ABQ37429;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 24020.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP010074.
XX
PR 01-SEP-2000; 2000DE-01043826.
XX
PR 03-SEP-2000; 2000DE-01044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful for
PT diagnosis and prognosis, comprises selective hybridization of amplicons
PT from chemically treated DNA.
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC investigating cell differentiation. The method allows the methylation
CC status of many C residues to be determined simultaneously. ABQ13410-
CC ABQ54121 represent genomic DNA sequences used to illustrate the method
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention.

QY Sequence 720 BP; 288 A; 268 C; 75 G; 89 T; 0 U; 0 Other;
XX

Query Match 86.3%; Score 16.4; DB 6; Length 720;
Best Local Similarity 94.4%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCTTCGCGAAAAAATCG 19

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCTTCGCGAAAAAATCG 19
DB 580 CCTTCGCGAAAAAATCG 597

RESULT 5
ABQ37428/c
ID ABQ37428 standard; DNA; 720 BP.
XX
AC ABQ37428;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 24019.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP010074.
XX
PR 01-SEP-2000; 2000DE-01043826.
XX
PR 05-SEP-2000; 2000DE-01044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful for
PT diagnosis and prognosis, comprises selective hybridization of amplicons
PT from chemically treated DNA.
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC investigating cell differentiation. The method allows the methylation
CC status of many C residues to be determined simultaneously. ABQ13410-
CC ABQ54121 represent genomic DNA sequences used to illustrate the method
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention.

QY Sequence 720 BP; 89 A; 75 C; 268 G; 288 T; 0 U; 0 Other;
XX

Query Match 86.3%; Score 16.4; DB 6; Length 720;
Best Local Similarity 94.4%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCTTCGCGAAAAAATCG 19

XX 27-AUG-2003 (revised)
DT 25-MAR-2003 (revised)
DT 18-FEB-1993 (first entry)
XX
XX AcMNPV gp67 signal sequence.
DE
XX Toxin; Butus eupaeus; baculovirus; AcMNPV; insecticide; scorpion;
KW A. californica nuclear polyhedrosis virus; juvenile hormone esterase; ss.
KW
XX Autographa californica nucleopolyhedrovirus.
OS
XX EP505207-A1.
PN
XX 23-SEP-1992.
PD
XX 20-MAR-1992; 92EP-00302439.
PF
XX 22-MAR-1991; 91GB-00006185.
PR
XX (WELL) WELLCOME FOUND LTD.
PA (NATU-) NATURAL ENVIRONMENT RES COUNCIL.
PA (ROUS) ROUSSEL-UCIAP.
PA (HMRI) HOECHST SCHERING AGREVO SA.
XX
XX Cayley PJ, Stewart LMD, Possee RD, Ferber ML;
PI WPI; 1992-318085/39.
XX P-PSDB; AAR27265.
DR
XX Recombinant baculovirus used as insect control agent - causes infected
PT insect cells to secrete toxin e.g. spider or scorpion.
XX
XX Disclosure; Fig 1; 45pp; English.
PS
XX The sequences given in AA028603 and AA028610-14 encode signal peptides
CC which were used to allow the secretion of a toxin from the cells in which
CC they were produced. The toxin used was an insect specific scorpion toxin
CC from Butus eupaeus and it was expressed under the control of a
CC baculovirus promoter in a recombinant A. californica Nuclear polyhedrosis
CC virus (AcMNPV). Baculoviruses are host selective and can be used for
CC insecticides as they are non-polluting to the environment. Suitable
CC signal sequences for the secretion of the toxin into the insect host may
CC be derived from mammalian or insect cells. These signal peptides are
CC cleaved from the toxin so that the toxin can be secreted. The signal
CC peptides can therefore be fused to either the N-terminal or C-terminal
CC amino acid residue of the toxin either directly or via a linker sequence.
CC (Updated on 25-MAR-2003 to correct PN field.) (Updated on 25-MAR-2003 to
CC correct PA field.) (Updated on 27-AUG-2003 to correct OS field.)
XX
XX Sequence 327 BP; 104 A; 72 C; 67 G; 84 T; 0 U; 0 Other;
SQ
Query Match 83.2%; Score 15.8; DB 2; Length 327;
Best Local Similarity 89.5%; Pred. No. 2.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GCCTTCGCGAAAAAATCG 19
DB 106 GCCTTCGCGAAAAAATCG 124
RESULT 9
ABZ52684/C
ID ABZ52684 standard; cDNA; 369 BP.
XX
XX ABZ52684;
AC
XX 28-MAR-2003 (first entry)
DT
XX Aspergillus oryzae polynucleotide SEQ ID NO 1797.
DE
XX Aspergillus oryzae; fermentation; fungus; industrial; EST;
KW expressed sequence tag; gene; ss.

XX Aspergillus oryzae.
OS
XX WO200279476-A1.
PN
XX 10-OCT-2002.
PD
XX 22-MAR-2002; 2002WO-IB000890.
PF
XX 30-MAR-2001; 2001JP-00098371.
PR
XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA (NARE-) NAT RES INST BREWING.
PA (NORQ) NAT FOOD RES INST MIN AGRIC.
XX
XX Machida M, Akita O, Kashiwagi Y, Kitamoto K, Horiuchi H;
PI Takeuchi M, Kobayashi T, Kitamoto N, Gomi K, Abe K;
XX WPI; 2003-046817/04.
DR
XX Detection of expression of specific Aspergillus genes for monitoring the
PT fermentation and growth conditions of the fungus, using DNA probes.
XX
XX Claim 1; SEQ ID NO 1797; 48pp + Sequence Listing; Japanese.
PS
XX The invention relates to a polynucleotide having any of 6006 specific
CC sequences (ABZ50888-ABZ56893), which are expressed by a fungus under
CC specific culture conditions including one or more of eutrophic,
CC oligotrophic, solid, early germination, alkaline, high temperature, low
CC temperature or maltose culture or polynucleotides stringently hybridising
CC to these sequences. The polynucleotides are useful for monitoring the
CC progress of fermentation and the growth conditions of a fungus,
CC especially of Aspergillus oryzae which is widely used in industrial
CC fermentation. Also monitoring for fungal contamination. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 369 BP; 97 A; 82 C; 82 G; 107 T; 0 U; 1 Other;
SQ
Query Match 83.2%; Score 15.8; DB 7; Length 369;
Best Local Similarity 89.5%; Pred. No. 2.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GCCTTCGCGAAAAAATCG 19
DB 118 GCCTTCGCGAAAAAATCG 100
RESULT 10
ABQ47425
ID ABQ47425 standard; DNA; 522 BP.
XX
XX ABQ47425;
AC
XX 12-JUL-2002 (first entry)
DT
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 34016.
DE
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
XX Homo sapiens.
OS
XX WO200218632-A2.
PN
XX 07-MAR-2002.
PD
XX 01-SEP-2001; 2001WO-EP010074.
PF
XX 01-SEP-2000; 2000DE-01043826.
PR

DR WPI; 2002-371829/40.

XX Determining the degree of cytosine methylation in genomic DNA, useful for

PT diagnosis and prognosis, comprises selective hybridization of amplicons

PT from chemically treated DNA.

PS Claim 12; 56pp + Sequence Listing; 56pp; German.

XX This invention describes a novel method for determining the degree of

CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a

CC genomic sample of DNA. The sample is treated chemically to convert

CC cytosine (C) but not methylated C, to uracil, then part of the genomic

CC DNA that contains the target C is amplified to form a labeled amplicon.

CC The amplicon is hybridised to two classes, each with at least one member,

CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the

CC degree of hybridisation to both classes is determined from the label on

CC the amplicon. From the ratio of labels hybridised to the two classes of

CC oligomers, the degree of methylation is calculated. The method is used:

CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs

CC and of a wide range of diseases, e.g. cancer, disorders of the central

CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,

CC particularly by detecting mutations or single nucleotide polymorphisms

CC (SNP's); and (ii) for differentiation of cell or tissue types and for

CC investigating cell differentiation. The method allows the methylation

CC status of many C residues to be determined simultaneously. ABQ13410-

CC ABQ54121 represent genomic DNA sequences used to illustrate the method

CC for determining the degree of cytosine methylation described in the

CC disclosure of the invention

XX

SQ Sequence 524 BP; 99 A; 54 C; 187 G; 184 T; 0 U; 0 Other;

Query Match 83.2%; Score 15.8; DB 6; Length 524;

Best Local Similarity 89.5%; Pred. No. 2.8e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCCTTCGCGAAAAAATCG 19

Db 173 GCCTTAACGAAAAAATCG 155

RESULT 13

ABQ47199

ID ABQ47199 standard; DNA; 524 BP.

XX AC ABQ47199;

XX DT 12-JUL-2002 (first entry)

XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 33790.

XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;

KW drug; side effect; cancer; central nervous system; cardiovascular;

KW gastrointestinal; respiratory system; single nucleotide polymorphism;

KW SNP; cell differentiation; ds.

XX OS Homo sapiens.

XX PN WO200218632-A2.

XX PD 07-MAR-2002.

XX PF 01-SEP-2001; 2001WO-EP010074.

XX PR 01-SEP-2000; 2000DE-01043826.

PR 05-SEP-2000; 2000DE-01044543.

XX PA (EPIG-) EPIGENOMICS AG.

XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;

XX WPI; 2002-371829/40.

XX Determining the degree of cytosine methylation in genomic DNA, useful for

PT diagnosis and prognosis, comprises selective hybridization of amplicons

PT from chemically treated DNA.

PT diagnosis and prognosis, comprises selective hybridization of amplicons

PT from chemically treated DNA.

PS Claim 12; 56pp + Sequence Listing; 56pp; German.

XX This invention describes a novel method for determining the degree of

CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a

CC genomic sample of DNA. The sample is treated chemically to convert

CC cytosine (C) but not methylated C, to uracil, then part of the genomic

CC DNA that contains the target C is amplified to form a labeled amplicon.

CC The amplicon is hybridised to two classes, each with at least one member,

CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the

CC degree of hybridisation to both classes is determined from the label on

CC the amplicon. From the ratio of labels hybridised to the two classes of

CC oligomers, the degree of methylation is calculated. The method is used:

CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs

CC and of a wide range of diseases, e.g. cancer, disorders of the central

CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,

CC particularly by detecting mutations or single nucleotide polymorphisms

CC (SNP's); and (ii) for differentiation of cell or tissue types and for

CC investigating cell differentiation. The method allows the methylation

CC status of many C residues to be determined simultaneously. ABQ13410-

CC ABQ54121 represent genomic DNA sequences used to illustrate the method

CC for determining the degree of cytosine methylation described in the

CC disclosure of the invention

XX

SQ Sequence 524 BP; 184 A; 187 C; 54 G; 99 T; 0 U; 0 Other;

Query Match 83.2%; Score 15.8; DB 6; Length 524;

Best Local Similarity 89.5%; Pred. No. 2.8e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCCTTCGCGAAAAAATCG 19

Db 352 GCCTTAACGAAAAAATCG 370

RESULT 14

ABQ27793

ID ABQ27793 standard; DNA; 600 BP.

XX AC ABQ27793;

XX DT 12-JUL-2002 (first entry)

XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 14384.

XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;

KW drug; side effect; cancer; central nervous system; cardiovascular;

KW gastrointestinal; respiratory system; single nucleotide polymorphism;

KW SNP; cell differentiation; ds.

XX OS Homo sapiens.

XX PN WO200218632-A2.

XX PD 07-MAR-2002.

XX PF 01-SEP-2001; 2001WO-EP010074.

XX PR 01-SEP-2000; 2000DE-01043826.

PR 05-SEP-2000; 2000DE-01044543.

XX PA (EPIG-) EPIGENOMICS AG.

XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;

XX WPI; 2002-371829/40.

XX Determining the degree of cytosine methylation in genomic DNA, useful for

PT diagnosis and prognosis, comprises selective hybridization of amplicons

PT from chemically treated DNA.

PS Claim 12; 56pp + Sequence Listing; 56pp; German.

XX This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one member,
 CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
 CC degree of hybridisation to both classes is determined from the label on
 CC the amplicon. From the ratio of labels hybridised to the two classes of
 CC oligomers, the degree of methylation is calculated. The method is used:
 CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
 CC and of a wide range of diseases, e.g. cancer, disorders of the central
 CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
 CC particularly by detecting mutations or single nucleotide polymorphisms
 CC (SNP's); and (ii) for differentiation of cell or tissue types and for
 CC investigating cell differentiation. The method allows the methylation
 CC status of many C residues to be determined simultaneously. ABQ13410-
 CC ABQ54121 represent genomic DNA sequences used to illustrate the method
 CC for determining the degree of cytosine methylation described in the
 CC disclosure of the invention

SQ Sequence 600 BP; 255 A; 186 C; 73 G; 86 T; 0 U; 0 Other;

Query Match 83.2%; Score 15.8; DB 6; Length 600;
 Best Local Similarity 89.5%; Pred. No. 2.8e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCCTTCGCGAAAAAATCG 19
 |||||
 DB 113 GCCTACGCGAAAAAACC 131

RESULT 15

ABQ27792/c
 ID ABQ27792 standard; DNA; 600 BP.

XX AC ABQ27792;

XX 12-JUL-2002 (first entry)

XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 14383.

XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KW drug; side effect; cancer; central nervous system; cardiovascular;
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;
 KW SNP; cell differentiation; ds.

XX Homo sapiens.

XX WO200218632-A2.

XX 07-MAR-2002.

XX 01-SEP-2001; 2001WO-EP010074.

XX 01-SEP-2000; 2000DE-01043826.

XX 05-SEP-2000; 2000DE-01044543.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K, Guetig D;

XX WPI; 2002-371829/40.

XX Determining the degree of cytosine methylation in genomic DNA, useful for
 PT diagnosis and prognosis, comprises selective hybridization of amplicons
 PT from chemically treated DNA.

XX Claim 12; 56pp + Sequence Listing; 56pp; German.

XX This invention describes a novel method for determining the degree of

CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one member,
 CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
 CC degree of hybridisation to both classes is determined from the label on
 CC the amplicon. From the ratio of labels hybridised to the two classes of
 CC oligomers, the degree of methylation is calculated. The method is used:
 CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
 CC and of a wide range of diseases, e.g. cancer, disorders of the central
 CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
 CC particularly by detecting mutations or single nucleotide polymorphisms
 CC (SNP's); and (ii) for differentiation of cell or tissue types and for
 CC investigating cell differentiation. The method allows the methylation
 CC status of many C residues to be determined simultaneously. ABQ13410-
 CC ABQ54121 represent genomic DNA sequences used to illustrate the method
 CC for determining the degree of cytosine methylation described in the
 CC disclosure of the invention

XX SQ Sequence 600 BP; 86 A; 73 C; 186 G; 255 T; 0 U; 0 Other;

Query Match 83.2%; Score 15.8; DB 6; Length 600;
 Best Local Similarity 89.5%; Pred. No. 2.8e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCCTTCGCGAAAAAATCG 19
 |||||
 DB 488 GCCTACGCGAAAAAACC 470

Search completed: April 24, 2004, 18:45:07
 Job time : 162.579 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2004, 18:26:15 ; Search time 35.1852 Seconds
(without alignments)
299.673 Million cell updates/sec

Title: US-10-084-555A-118

Perfect score: 19

Sequence: 1 gccttcgcgaataatcg 19

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*

2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*

3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*

4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*

5: /cgn2_6/ptodata/2/ina/PTUS COMB.seq.*

6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15.8	83.2	30	1	US-08-451-472-37
2	15.8	83.2	327	1	US-08-451-472-22
3	15.8	83.2	578	1	US-08-451-472-32
C 4	15.8	83.2	1422	4	US-09-489-039A-1817
C 5	15.8	83.2	48908	4	US-09-453-702B-137
6	15.4	81.1	534	4	US-09-252-991A-4043
C 7	15.4	81.1	5552	4	US-09-676-974-2
8	15.4	81.1	5552	4	US-09-554-724B-4
9	15	78.9	1740	4	US-09-543-681A-697
10	14.8	77.9	6066	4	US-09-543-681A-1912
C 11	14.8	77.9	640681	4	US-09-790-988-1
12	14.4	75.8	618	4	US-09-107-532A-2867
13	14.4	75.8	3190	4	US-09-221-017B-314
C 14	14.4	75.8	4360	1	US-08-486-380-1
C 15	14.4	75.8	4360	2	US-08-436-080-1
C 16	14.4	75.8	4360	2	US-08-250-848-1
C 17	14.4	75.8	4360	2	US-08-438-192-1
C 18	14.4	75.8	4360	2	US-08-475-971-1
C 19	14.4	75.8	4360	3	US-09-021-203-1
C 20	14.4	75.8	4360	4	US-08-122-352-1
C 21	14.4	75.8	4360	4	US-08-122-090-1
C 22	14.4	75.8	4360	4	US-08-480-186-1
C 23	14.4	75.8	4360	4	US-08-487-495-1
C 24	14.4	75.8	5293	4	US-10-204-708-51
25	14.4	75.8	8022	4	US-09-029-047C-3
26	14.2	74.7	225	4	US-09-107-532A-2192
27	14.2	74.7	456	4	US-09-621-976-2178

28	14.2	74.7	581	4	US-09-669-751-58	Sequence 58, Appl
C 29	14.2	74.7	660	4	US-09-252-991A-13837	Sequence 13837, A
30	14.2	74.7	699	4	US-09-252-991A-575	Sequence 575, App
31	14.2	74.7	738	4	US-08-282-991A-13412	Sequence 13412, A
32	14.2	74.7	771	4	US-09-252-991A-13618	Sequence 13618, A
33	14.2	74.7	810	4	US-08-724-623-11	Sequence 11, Appl
34	14.2	74.7	923	1	US-08-376-296-9	Sequence 9, Appl
35	14.2	74.7	923	3	US-08-910-691-10	Sequence 10, Appl
36	14.2	74.7	969	1	US-08-376-296-6	Sequence 6, Appl
37	14.2	74.7	969	3	US-08-910-691-6	Sequence 6, Appl
38	14.2	74.7	1245	4	US-09-252-991A-8143	Sequence 8143, Ap
39	14.2	74.7	1310	4	US-08-766-534-526	Sequence 526, App
40	14.2	74.7	1410	4	US-09-252-991A-6871	Sequence 6871, Ap
C 41	14.2	74.7	1434	4	US-09-252-991A-522	Sequence 522, App
42	14.2	74.7	1458	4	US-09-252-991A-599	Sequence 599, App
C 43	14.2	74.7	1482	4	US-09-252-991A-6850	Sequence 6850, Ap
C 44	14.2	74.7	1500	4	US-09-252-991A-6757	Sequence 6757, Ap
45	14.2	74.7	1681	4	US-09-634-238-51	Sequence 51, Appl

ALIGNMENTS

RESULT 1

US-08-451-472-37

; Sequence 37, Application US/08451472

; Patent No. 5770192

; GENERAL INFORMATION:

; APPLICANT:

; TITLE OF INVENTION: BIOLOGICAL CONTROL AGENTS

; NUMBER OF SEQUENCES: 73

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Nixon & Vanderhye PC

; STREET: 8th Floor, 1100 No. 5770192th Glebe Road

; CITY: Arlington

; STATE: Virginia

; COUNTRY: USA

; ZIP: 22201-4714

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/451.472

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/117,125

; FILING DATE: 24-NOV-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Mary J Wilson

; REGISTRATION NUMBER: 32,955

; REFERENCE/DOCKET NUMBER: 117-187

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 816 4000

; TELEFAX: (703) 816 4100

; TELEX: 200797 NIXN UR

; INFORMATION FOR SEQ ID NO: 37:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 30 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: CDNA

US-08-451-472-37

Query Match 83.2%; Score 15.8; DB 1; Length 30;

Best Local Similarity 89.5%; Pred. No. 21;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCCTTCGCGAAAAATCG 19

||||| ||||||||| ||

Db 7 GCCTTGGCGAAAAAATCG 25

RESULT 2

US-08-451-472-22
; Sequence 22, Application US/08451472
; Patent No. 5770192
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: BIOLOGICAL CONTROL AGENTS
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon & Vanderhye PC
; STREET: 8th Floor, 1100 No. 5770192th Glebe Road
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; APPLICATION DATA:
; APPLICATION NUMBER: US/08/451.472
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; REFERENCE/DOCKET NUMBER: 117-187
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816 4000
; TELEFAX: (703) 816 4100
; TELEX: 200797 NIXN UR
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 327 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..327
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary J Wilson
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 117-187
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816 4000
; TELEFAX: (703) 816 4100
; TELEX: 200797 NIXN UR
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 327 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..327

US-08-451-472-22
Query Match 83.2%; Score 15.8; DB 1; Length 327;
Best Local Similarity 89.5%; Pred. No. 30;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCCTTGGCGAAAAAATCG 19

Db 106 GCCTTGGCGAAAAAATCG 124

RESULT 3

US-08-451-472-32
; Sequence 32, Application US/08451472
; Patent No. 5770192
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: BIOLOGICAL CONTROL AGENTS
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon & Vanderhye PC
; STREET: 8th Floor, 1100 No. 5770192th Glebe Road
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; APPLICATION DATA:
; APPLICATION NUMBER: US/08/451.472
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; REFERENCE/DOCKET NUMBER: 117-187
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816 4000
; TELEFAX: (703) 816 4100
; TELEX: 200797 NIXN UR
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 327 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..327
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary J Wilson
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 117-187
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816 4000
; TELEFAX: (703) 816 4100
; TELEX: 200797 NIXN UR
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 327 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..327

US-08-451-472-32
Query Match 83.2%; Score 15.8; DB 1; Length 327;
Best Local Similarity 89.5%; Pred. No. 30;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCCTTGGCGAAAAAATCG 19

Db 106 GCCTTGGCGAAAAAATCG 124

US-08-451-472-32
; Sequence 32, Application US/08451472
; Patent No. 5770192
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: BIOLOGICAL CONTROL AGENTS
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon & Vanderhye PC
; STREET: 8th Floor, 1100 No. 5770192th Glebe Road
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; APPLICATION DATA:
; APPLICATION NUMBER: US/08/451.472
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; REFERENCE/DOCKET NUMBER: 117-187
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816 4000
; TELEFAX: (703) 816 4100
; TELEX: 200797 NIXN UR
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 578 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 246..569
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary J Wilson
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 117-187
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816 4000
; TELEFAX: (703) 816 4100
; TELEX: 200797 NIXN UR
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 578 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 246..569

US-08-451-472-32
Query Match 83.2%; Score 15.8; DB 1; Length 578;
Best Local Similarity 89.5%; Pred. No. 33;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCCTTGGCGAAAAAATCG 19

Db 351 GCCTTGGCGAAAAAATCG 369

RESULT 4

US-09-489-039A-1817/c
; Sequence 1817, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709-2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 1817
; LENGTH: 1422
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-1817

US-09-489-039A-1817/c
Query Match 83.2%; Score 15.8; DB 4; Length 1422;
Best Local Similarity 89.5%; Pred. No. 38;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCCTTGGCGAAAAAATCG 19

Db 1086 GCCTTGGCGAAAAAATCG 1068

RESULT 5

US-09-453-702B-137/c

```
; Sequence 137, Application US/09453702B
; Patent No. 6365723
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; Burland, Valerie
; Perna, Nicole T.
; Plunkett, Guy
; Welch, Rod
; TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/453.702B
; FILING DATE: 03-Dec-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 137:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 48908
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 137:
US-09-453-702B-137

Query Match      83.2%; Score 15.8; DB 4; Length 48908;
Best Local Similarity 89.5%; Pred. No. 64;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 GCCTTCGCGAAAAAATCG 19
Db      3585 GCCTTCGCGAAAAAATCG 3567

RESULT 6
US-09-252-991A-4043
; Sequence 4043, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4043
; LENGTH: 534
```

```
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4043

Query Match      81.1%; Score 15.4; DB 4; Length 534;
Best Local Similarity 94.1%; Pred. No. 53;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      3 CTTTCGCGAAAAAATCG 19
Db      91 CTTTCGCGAAAAAATCG 107

RESULT 7
US-09-676-974-2/c
; Sequence 2, Application US/09676974
; Patent No. 6391631
; GENERAL INFORMATION:
; APPLICANT: HACKER, JORG
; APPLICANT: SONNEN-BORN, ULRICH
; APPLICANT: SCHULZE JURGEN
; APPLICANT: BLUM-OEHLER, GABRIELE
; APPLICANT: MALINKA, JURGEN
; APPLICANT: PROPERT, HANS
; TITLE OF INVENTION: BACTERIAL PLASMIDS
; FILE REFERENCE: 11347/268416/BET
; CURRENT APPLICATION NUMBER: US/09/676,974
; CURRENT FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: PCT/EP98/01720
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: DE 197 13543.9
; PRIOR FILING DATE: 1997-04-02
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 5552
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (120)
; OTHER INFORMATION: A, T, C or G
; NAME/KEY: modified_base
; LOCATION: (661)
; OTHER INFORMATION: A, T, C or G
; NAME/KEY: modified_base
; LOCATION: (1290)
; OTHER INFORMATION: A, T, C or G
; NAME/KEY: modified_base
; LOCATION: (5341)
; OTHER INFORMATION: A, T, C or G
; NAME/KEY: modified_base
; LOCATION: (5400)
; OTHER INFORMATION: A, T, C or G
US-09-676-974-2

Query Match      81.1%; Score 15.4; DB 4; Length 5552;
Best Local Similarity 94.1%; Pred. No. 75;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2 CTTTCGCGAAAAAATCG 18
Db      3147 CTTTCGCGAAAAAATCG 3131

RESULT 8
US-09-554-724B-4/c
; Sequence 4, Application US/09554724B
; Patent No. 6489107
; GENERAL INFORMATION:
; APPLICANT: HACKER, JORG
; APPLICANT: SONNEN-BORN, ULRICH
; APPLICANT: BLUM-OEHLER, GABRIELE
```



```

CORRESPONDENCE ADDRESS:
ADDRESS: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354

COMPUTER READABLE FORM:
MEDIUM TYPE: CD/ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: JULY 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneka
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 2867:
SEQUENCE CHARACTERISTICS:
LENGTH: 618 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...618
SEQUENCE DESCRIPTION: SEQ ID NO: 2867:
US-09-107-532A-2867

```

```

Query Match 75.8%; Score 14.4; DB 4; Length 618;
Best Local Similarity 93.8%; Pred. No. 1.8e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 3 CTTCCGCGAAAAATC 18
Db 329 CTTCCGCGAAAAATC 344

```

```

RESULT 13
US-09-221-017B-314
Sequence 314, Application US/09221017B
Patent No. 6444799
GENERAL INFORMATION:
APPLICANT: Ross, Bruce C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
NUMBER OF SEQUENCES: 1120
CORRESPONDENCE ADDRESS:
ADDRESS: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FastSEQ for Windows Version 2.0b

```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221,017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PPI182
FILING DATE: 31-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PPI546
FILING DATE: 30-JAN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP2911
FILING DATE: 09-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU98/01023
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Moroy, Gladys H.
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 314:
SEQUENCE CHARACTERISTICS:
LENGTH: 3190 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: PORYPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc feature
LOCATION: 1...3190
US-09-221-017B-314

```

```

Query Match 75.8%; Score 14.4; DB 4; Length 3190;
Best Local Similarity 93.8%; Pred. No. 2.3e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 4 TTCCGCGAAAAAATCG 19
Db 2061 TTCCGCGAAAAAATCG 2076

```

```

RESULT 14
US-08-486-380-1/c
Sequence 1, Application US/08486380
Patent No. 5581036
GENERAL INFORMATION:
APPLICANT: Rangan, Thirumale S.
APPLICANT: Rajasekaran, Kanniah
TITLE OF INVENTION: Regeneration of Cotton Plants in
TITLE OF INVENTION: Suspension Culture
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESS: Christie, Parker & Hale
STREET: P.O. Box 7068
CITY: Pasadena
STATE: CA
COUNTRY: USA
ZIP: 91109-7068
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/486,380

```

```
;
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/122,351
; FILING DATE:
; APPLICATION NUMBER: US 07/122,200
; FILING DATE: 18-NOV-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/680,048
; FILING DATE: 29-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharp Esq., Janice A.
; REGISTRATION NUMBER: 34,051
; REFERENCE/DOCKET NUMBER: P114:25601
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (818)795-5843
; TELEFAX: (818)795-5843
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4360 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Bacillus thuringiensis
;
; US-08-486-380-1
;
; Query Match 75.8%; Score 14.4; DB 1; Length 4360;
; Best Local Similarity 93.8%; Pred. No. 2.4e+02;
; Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
;
; Qy 1 GCCTTCGCGAAAAAA 16
; Db 3794 GCCTTCGCGAAAAAA 3779
;
; RESULT 15
; US-08-436-080-1/c
; Sequence 1, Application US/08436080
; Patent No. 5834292
; GENERAL INFORMATION:
; APPLICANT: Rangan, Thirumale S.
; APPLICANT: Anderson, David M.
; TITLE OF INVENTION: Method for Producing Somaclonal Variant
; TITLE OF INVENTION: Cotton Plants
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christie, Parker & Hale
; STREET: P.O. Box 7068
; CITY: Pasadena
; STATE: CA
; COUNTRY: USA
; ZIP: 91109-7068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/436,080
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/122,094
; FILING DATE:
; APPLICATION NUMBER: US 07/122,200
; FILING DATE: 18-NOV-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/680,048
; FILING DATE: 29-MAR-1991
```

```
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharp Esq., Janice A.
; REGISTRATION NUMBER: 34,051
; REFERENCE/DOCKET NUMBER: P114:25705
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (818)795-5843
; TELEFAX: (818)577-1769
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4360 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Bacillus thuringiensis
;
; US-08-436-080-1
;
; Query Match 75.8%; Score 14.4; DB 2; Length 4360;
; Best Local Similarity 93.8%; Pred. No. 2.4e+02;
; Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
;
; Qy 1 GCCTTCGCGAAAAAA 16
; Db 3794 GCCTTCGCGAAAAAA 3779
;
; Search completed: April 24, 2004, 21:13:49
; Job time : 36.1852 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2004, 19:25:41 ; Search time 146.84 Seconds
(without alignments)
583.385 Million cell updates/sec

Title: US-10-084-555A-118

Perfect score: 19
Sequence: 1 gccttcgcaaaaatcg 19

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2907579 seqs, 2254313464 residues

Total number of hits satisfying chosen parameters: 5815158

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:
- 3: /cgn2_6/ptodata/2/pubpna/US05_NEW_PUB.seq:
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:
- 8: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
- 13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
- 14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:
- 15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:
- 16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:
- 17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:
- 18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:
- 19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	15.8	83.2	134	9	US-09-560-863-936
2	15.8	83.2	1170	16	US-10-369-493-34351
3	15.8	83.2	1361	16	US-10-369-493-37099
4	15.8	83.2	2094	16	US-10-369-493-45478
5	15.8	83.2	3234	9	US-09-374-300-2168
6	15.8	83.2	4501	15	US-10-172-086-101
7	15.8	83.2	8245	13	US-10-221-714A-169
8	15.8	83.2	48908	15	US-10-114-170-137
9	15.4	81.1	489	13	US-10-389-647-264
10	15.4	81.1	831	13	US-10-027-632-161560
11	15.4	81.1	831	16	US-10-027-632-161560
12	15.4	81.1	1192	13	US-10-425-114-4729
13	15.4	81.1	1515	13	US-10-282-122A-40479
14	15.4	81.1	1558	17	US-10-333-946-30

C 15	15.4	81.1	2000	9	US-09-938-842A-3490	Sequence 3490, Ap
C 16	15.4	81.1	2000	11	US-09-938-842A-3490	Sequence 3490, Ap
C 17	15.4	81.1	6099	13	US-10-221-613-308	Sequence 308, App
C 18	15.4	81.1	213251	16	US-10-398-221-6	Sequence 6, Appli
C 19	15.4	81.1	3011208	16	US-10-398-221-2058	Sequence 2058, Ap
C 20	15	78.9	354	9	US-09-974-300-4339	Sequence 4339, Ap
C 21	15	78.9	507	13	US-10-282-122A-13897	Sequence 13897, A
C 22	15	78.9	1728	13	US-10-282-122A-32258	Sequence 32258, A
C 23	15	78.9	7278	16	US-10-398-221-3819	Sequence 3819, Ap
C 24	15	78.9	11178	13	US-10-221-613-15	Sequence 15, Appl
C 25	15	78.9	1163020	16	US-10-398-221-10	Sequence 10, Appl
C 26	14.8	77.9	163	13	US-10-424-559-47542	Sequence 47542, A
C 27	14.8	77.9	256	9	US-09-974-300-1987	Sequence 1987, Ap
C 28	14.8	77.9	278	13	US-10-424-559-35806	Sequence 35806, A
C 29	14.8	77.9	335	13	US-10-424-559-29370	Sequence 29370, A
C 30	14.8	77.9	480	13	US-10-424-559-82542	Sequence 82542, A
C 31	14.8	77.9	506	13	US-10-425-114-34195	Sequence 34195, A
C 32	14.8	77.9	564	13	US-10-027-632-252636	Sequence 252636,
C 33	14.8	77.9	564	16	US-10-027-632-252636	Sequence 252636,
C 34	14.8	77.9	999	13	US-10-335-977-2599	Sequence 2599, Ap
C 35	14.8	77.9	1113	13	US-10-335-977-2600	Sequence 2600, Ap
C 36	14.8	77.9	1173	16	US-10-369-493-40925	Sequence 40925, A
C 37	14.8	77.9	1389	13	US-10-282-122A-42328	Sequence 42328, A
C 38	14.8	77.9	1443	13	US-10-282-122A-24184	Sequence 24184, A
C 39	14.8	77.9	1542	16	US-10-369-493-45820	Sequence 45820, A
C 40	14.8	77.9	1602	9	US-09-801-368-161	Sequence 161, App
C 41	14.8	77.9	2270	15	US-10-101-484A-460	Sequence 460, App
C 42	14.8	77.9	2713	16	US-10-369-493-27089	Sequence 27089, A
C 43	14.8	77.9	5442	15	US-10-311-455-1942	Sequence 1942, Ap
C 44	14.8	77.9	5972	15	US-10-251-364-9	Sequence 9, Appli
C 45	14.8	77.9	5972	15	US-10-114-739A-69	Sequence 69, Appl

ALIGNMENTS

RESULT 1

US-09-560-863-936
; Sequence 936, Application US/09560863
; Patent No. US20020110809A1
; GENERAL INFORMATION:
; APPLICANT: Nehls, Michael C.
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20020110809A1 Human Polynucleotides and the
; FILE OF INVENTION: Polypeptides Encoded Thereby
; FILE REFERENCE: LEX-0018-USA
; CURRENT APPLICATION NUMBER: US/09/560,863
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/132,408
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 1008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 936
; LENGTH: 134
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-560-863-936

Query Match 83.2%; Score 15.8; DB 9; Length 134;
Best Local Similarity 89.5%; Pred. No. 4.1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCCTTCGCAAAAATCG 19
DB 53 GCCTACGTGAAAAATCG 71

RESULT 2
US-10-369-493-34351
; Sequence 34351, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:

```
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; LENGTH: 1170
; TYPE: DNA
; ORGANISM: Sphingomonas aromaticivorans
US-10-369-493-34351

Query Match      83.2%; Score 15.8; DB 16; Length 1170;
Best Local Similarity 89.5%; Pred. No. 5.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCCTTCGCGAAAAAATCG 19
Db 577 GCCTTCGCGAACGAAATCG 595

RESULT 3
US-10-369-493-37099
; Sequence 37099, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; LENGTH: 1361
; TYPE: DNA
; ORGANISM: Aspergillus nidulans
US-10-369-493-37099

Query Match      83.2%; Score 15.8; DB 16; Length 1361;
Best Local Similarity 89.5%; Pred. No. 5.3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCCTTCGCGAAAAAATCG 19
Db 34 GCCTTCGCGAACACAGTCG 52

RESULT 4
US-10-369-493-46478
; Sequence 46478, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
```

```
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; LENGTH: 2094
; TYPE: DNA
; ORGANISM: Schizosaccharomyces pombe
US-10-369-493-46478

Query Match      83.2%; Score 15.8; DB 16; Length 2094;
Best Local Similarity 89.5%; Pred. No. 5.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCCTTCGCGAAAAAATCG 19
Db 297 GCCTTCGCGAAAAAATCG 315

RESULT 5
US-09-974-300-2168
; Sequence 2168, Application US/09974300
; Patent No. US20020148721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085-500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2168
; LENGTH: 3234
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-2168

Query Match      83.2%; Score 15.8; DB 9; Length 3234;
Best Local Similarity 89.5%; Pred. No. 5.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCCTTCGCGAAAAAATCG 19
Db 2403 GCCTTCGCGAAAAAATCG 2421

RESULT 6
US-10-172-086-101/c
; Sequence 101, Application US/10172086
; Publication No. US20030113750A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Method and nucleic acids for the differentiation
; TITLE OF INVENTION: of prostate tumors
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/172,086
; PRIOR FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 101
; LENGTH: 4501
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-172-086-101
```



```
Db          46 CTTTCGCGAAAAAATCG 62

RESULT 10
US-10-027-632-161560
; Sequence 161560, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 161560
; LENGTH: 831
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-161560

Query Match      81.1%; Score 15.4; DB 13; Length 831;
Best Local Similarity 94.1%; Pred. No. 7.9e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 CTTTCGCGAAAAAATC 18
        ||||| ||||| |||||
Db      654 CTTTCGCGAAAAAATC 670

RESULT 11
US-10-027-632-161560
; Sequence 161560, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 161560
; LENGTH: 831
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-161560

Query Match      81.1%; Score 15.4; DB 13; Length 831;
Best Local Similarity 94.1%; Pred. No. 7.9e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 CTTTCGCGAAAAAATC 18
        ||||| ||||| |||||
Db      654 CTTTCGCGAAAAAATC 670

RESULT 12
US-10-425-114-14729/c
; Sequence 14729, Application US/10425114
; Publication No. US20040034889A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Fihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 14729
; LENGTH: 1192
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB25-056-B4_FLI
US-10-425-114-14729

Query Match      81.1%; Score 15.4; DB 13; Length 1192;
Best Local Similarity 94.1%; Pred. No. 8.3e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 CTTTCGCGAAAAAATCG 19
        ||||| ||||| |||||
Db      46 CTTTCGCGAAAAAATCG 30

RESULT 13
US-10-282-122A-40479
; Sequence 40479, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haseibeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
```

```
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 40479
; LENGTH: 1515
; TYPE: DNA
; ORGANISM: Treponema pallidum
US-10-282-122A-40479

Query Match      81.1%; Score 15.4; DB 13; Length 1515;
Best Local Similarity 94.1%; Pred. No. 8.5e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 CCTTCGCGGAAAAAATC 18
Db      192 CCTTCGCGGAAAGAATC 208

RESULT 14
US-10-333-946-30/c
; Sequence 30, Application US/10333946
; Publication No. US20040023252A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.; THORNTON, Michael B.
; APPLICANT: ARVIZO, Chandra S.; LAL, Preeti G.
; APPLICANT: BURFORD, Neil; YUE, Henry
; APPLICANT: GANDHI, Ameena R.; ELLIOTT, Vicki S.
; APPLICANT: RAMKUMAR, Javalaxmi; BAUGHN, Mariah R.
; APPLICANT: KALLICK, Deborah A.; CHAVIA, Navinder K.
; APPLICANT: HAPALIA, April J.A.; YAO, Monique G.
; APPLICANT: LU, Yan; TRIBOULEY, Catherine M.
; APPLICANT: POLICKY, Jennifer L.; KEARNEY, Liam
; APPLICANT: GRAUL, Richard C.; WARREN, Bridget A.
; APPLICANT: LEE, Ernestine A.; DING, Li
; TITLE OF INVENTION: G-PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: PI-0176 USN
; CURRENT APPLICATION NUMBER: US/10/333,946
; CURRENT FILING DATE: 2003-01-22
; PRIOR APPLICATION NUMBER: PCT/US01/23433
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: US 60/221,478
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/223,268
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: US 60/227,054
; PRIOR FILING DATE: 2000-08-21
; PRIOR APPLICATION NUMBER: US 60/231,121
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: US 60/232,243
; PRIOR FILING DATE: 2000-09-13
; PRIOR APPLICATION NUMBER: US 60/232,691
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/235,146
; PRIOR FILING DATE: 2000-09-22
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PERL Program
; SEQ ID NO 30
```

```
; LENGTH: 1558
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20040023252A1 55002225CB1
US-10-333-946-30

Query Match      81.1%; Score 15.4; DB 17; Length 1558;
Best Local Similarity 94.1%; Pred. No. 8.5e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 GCCTTCGCGGAAAAAAT 17
Db      17 GCCTTCGCGGAAATAAT 1

RESULT 15
US-09-938-842A-3490/c
; Sequence 3490, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Krieps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 3490
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-3490

Query Match      81.1%; Score 15.4; DB 9; Length 2000;
Best Local Similarity 94.1%; Pred. No. 9.8e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 CTTCCGCGGAAAAAATCG 19
Db      1866 CGTCGCGGAAAAAATCG 1850

Search completed: April 24, 2004, 23:47:36
Job time : 149.84 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2004, 18:23:30 ; Search time 1478.95 Seconds
(without alignments)
383.638 Million cell updates/sec

Title: US-10-084-555A-118

Perfect score: 19

Sequence: 1 gccttcgcgaataatcg 19

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pin:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_pig:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	17.4	91.6	613	29	CG909674
C 2	17	89.5	363	12	B1711923
C 3	17	89.5	652	9	AL858455
C 4	17	89.5	895	28	AQ839817

C	5	17	89.5	1161	28	CC277174	CC277174	CH261-22G
	6	16.4	86.3	179	9	AV098279	AV098279	AV098279
	7	16.4	86.3	412	9	AL177223	AL177223	EST220834
C	8	16.4	86.3	452	13	BQ511872	BQ511872	EST1619287
	9	16.4	86.3	481	28	AZ049007	AZ049007	GSSB04040
C	10	16.4	86.3	531	10	BE343391	BE343391	EST408553
	11	16.4	86.3	551	10	BF496774	BF496774	AT106874.5
	12	16.4	86.3	599	28	CC356196	CC356196	PURQZ77TB
C	13	16.4	86.3	688	14	CK242884	CK242884	EST726521
	14	16.4	86.3	774	29	CC685856	CC685856	OGVEJ32TV
C	15	16.4	86.3	799	14	CK242883	CK242883	EST726520
	16	16.4	86.3	831	14	CK247403	CK247403	EST731040
C	17	16.4	86.3	861	28	CC356198	CC356198	PURQZ77TD
C	18	16.4	86.3	919	13	BQ144375	BQ144375	CH261-115
	19	16.4	86.3	1029	28	CC263236	CC263236	CF925968
C	20	16	84.2	509	29	CG777051	CG777051	1123004G0
	21	16	84.2	543	13	BM266492	BM266492	BM266492
	22	16	84.2	631	14	CD319296	CD319296	StrFus38.
C	23	16	84.2	746	12	BM338112	BM338112	MEST221-H
	24	16	84.2	818	9	AV869286	AV869286	AV869286
	25	16	84.2	822	29	CNS05Q11	CNS05Q11	Tetraodon
	26	16	84.2	841	13	BM266195	BM266195	BM266195
	27	16	84.2	882	9	AV991744	AV991744	AV991744
	28	16	84.2	885	9	AV837717	AV837717	AV837717
	29	16	84.2	889	13	BM279026	BM279026	BM279026
	30	16	84.2	908	13	BM255211	BM255211	BM255211
	31	16	84.2	918	9	AV846404	AV846404	AV846404
	32	16	84.2	1031	10	BE538631	BE538631	601064612
	33	16	84.2	1219	10	BE966032	BE966032	601659935
	34	15.8	83.2	111	9	AL177857	AL177857	EST221506
	35	15.8	83.2	149	10	BE987016	BE987016	UT-M-CG0P
C	36	15.8	83.2	169	12	BM736844	BM736844	952049G04
C	37	15.8	83.2	207	28	BH401996	BH401996	AG-ND-165
C	38	15.8	83.2	281	28	BH390619	BH390619	AG-ND-104
	39	15.8	83.2	312	12	BM737121	BM737121	952049G04
C	40	15.8	83.2	321	14	CF925968	CF925968	1af82d07.
C	41	15.8	83.2	409	12	B1511803	B1511803	B160007A
C	42	15.8	83.2	418	9	AL797459	AL797459	AL797459
	43	15.8	83.2	422	13	BU050284	BU050284	1111026D1
C	44	15.8	83.2	428	13	BE94652	BE94652	BE94652
	45	15.8	83.2	440	10	BF043630	BF043630	BF043630

ALIGNMENTS

RESULT 1
CG909674/c
LOCUS
DEFINITION
ZMMBB0314108.r ZMMBB Zea mays subsp. mays genomic clone
CG909674
CG909674
CG909674.1 GI:39608944
GSS.
Zea mays subsp. mays (maize)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 613)
Yu, Y., Kim, H.R., Hatfield, J., Soderlund, C., Bharti, A.K., Messing, J.
and Wing, R.
Sequencing of the maize genome
Unpublished (2003)
Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3367
Fax: 520 621 9288
Email: <http://genome.arizona.edu>
PCR Primers

```

FORWARD: T7
BACKWARD: M13r
Plate: 0314 row: I column: 08
Seq primer: M13r
Class: BAC ends.
Location/Qualifiers
1. .613
/organism="Zea mays subsp. mays"
/mol_type="genomic DNA"
/cultivar="B73"
/sub_species="mays"
/db_xref="taxon:4578"
/clone="ZMMB50314108"
/lab_host="DH10B"
/clone_lib="ZMMB50"
/note="vector: pBeloBAC11; Site 1: HindIII; Site 2:
HindIII; Zea mays L. sp. mays"
ORIGIN
Query Match 91.6%; Score 17.4; DB 29; Length 613;
Best Local Similarity 94.7%; Pred. No. 1.4e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCTTCGCGAAAAAATCG 19
|||||
Db 256 GCCTTCGCGAAAAAATCG 238

RESULT 2
BI7111923/c
LOCUS
DEFINITION ib76906.x4 Amplified Melton Mouse Islets 1 MIS1-A Mus musculus cDNA
clone IMAGE:5652682 3', similar to TR:004892 C04892 CYTOCHROME P450
LIKE TBP ;, mRNA sequence.
ACCESSION BI7111923.1 GI:15687618
VERSION BI7111923.1
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 363)
AUTHORS Melton,D., Meadows,A., Clifton,S., Hillier,L., Marra,M., Pape,D.,
Wyllie,T., Martin,J., Blistain,A., Schmitt,A., Theising,B.,
Ritter,E., Ronko,I., Bennett,J., Cardenas,M., Gibbons,M.,
McCann,R., Cole,R., Tsagarishvili,R., Williams,T., Jackson,Y. and
Bowers,Y.
WashU-Harvard Pancreas EST Project
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Juliana Brown
(brown@fas.harvard.edu)
MG1:1939008 This sequence now available from the IMAGE consortium,
for clone orders contact: info@image.llnl.gov
Seq primer: -40UP from Gibco.
Location/Qualifiers
1. .363
/organism="Mus musculus"
/mol_type="mRNA"
/strain="ICR"
/db_xref="taxon:10090"
/clone="IMAGE:5652682"
/sex="Male"
/tissue_type="Islets of Langerhans"

```

```

/dev_stage="Adult"
/lab_host="DH10B"
/clone_lib="Amplified Melton Mouse Islets 1 MIS1-A"
/note="Organ: Pancreas; Vector: pSPOR11; Site 1: Not 1;
Site 2: Sal 1; Library constructed using SuperScript
Plasmid Library kit (Life Technologies). cDNA made by
oligo-dT priming. Size-selected by column fractionation;
average insert size 0.91 kb. Amplified once on solid
support. cDNA Library Preparation: Guclon Chen."
ORIGIN
Query Match 89.5%; Score 17; DB 12; Length 363;
Best Local Similarity 100.0%; Pred. No. 2.3e+03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCTTCGCGAAAAAAT 17
|||||
Db 274 GCCTTCGCGAAAAAAT 258

RESULT 3
AL858455/c
LOCUS
DEFINITION AL858455 XGC-egg Silurana tropicalis cDNA clone TEG9059g14 5', mRNA
sequence.
ACCESSION AL858455
VERSION AL858455.2 GI:38633414
KEYWORDS EST.
SOURCE Silurana tropicalis (western clawed frog)
ORGANISM Silurana tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Silurana.
REFERENCE 1 (bases 1 to 652)
AUTHORS Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.
Sanger Xenopus tropicalis EST project 2001 (11_2003)
JOURNAL Unpublished (2003)
COMMENT On Sep 15, 2002 this sequence version replaced gi:22878636.
Contact: Taylor R
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS SEQUENCE ID: TEG9059g14.plkSP6
Sequencing primer: SP6
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn
cDNA was oligo dT primed from Sug of poly A+ RNA from egg.
EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the
5' end and NotI at the 3' end.
Vector: pCS107; Site 1: EcoRI; Site 2: NotI
Host: Escherichia coli XL1-blue.
Location/Qualifiers
1. .652
/organism="Silurana tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="TEG9059g14"
/dev_stage="egg"
/lab_host="Escherichia coli XL1-blue"
/clone_lib="XGC-egg"
/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
was oligo dT primed from Sug of poly A+ RNA from egg.
EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end"
FEATURES
source
ORIGIN
Query Match 89.5%; Score 17; DB 9; Length 652;
Best Local Similarity 100.0%; Pred. No. 2.1e+03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCTTCGCGAAAAAAT 17
|||||

```

Db 226 GCCTTCGCGAAAAAAT 210

RESULT 4
AQ839817

LOCUS
DEFINITION 895 bp DNA linear GSS 31-MAR-2000
260L13-C10 CITB Homo sapiens genomic clone 260L13, genomic survey
sequence.

ACCESSION
VERSION AQ839817 GI:6652449

KEYWORDS
SOURCE GSS.

ORGANISM
Homo sapiens (human)

REFERENCE
AUTHORS Cartpen, J.D., Makalowska, I., Robbins, C.M., Scott, N., Sood, R.,
Connors, T.D., Bonner, T.I., Smith, J.R., Faruque, M.U., Stephan, D.A.,
Pinkett, H., Morgenbesser, S.D., Su, K., Graham, C., Gregory, S.G.,
Williams, H., McDonald, L., Baxevasis, A.D., Klingler, K.W. and
Landes, G.W.

TITLE A 6-Mb high-resolution physical and transcription map encompassing
the hereditary prostate cancer 1 (HPC1) region

JOURNAL
MEDLINE Genomics 64 (1), 1-14 (2000)

PUBMED
10708513

COMMENT Contact: Cartpen JD
Cancer Genetics Branch
National Human Genome Research Institute/National Institutes of
Health
Bldg. 36, Room 3D04, 36 Convent Drive, Bethesda, MD
Tel: 301 435 5626
Fax: 301 435 5465
Email: jdc@ngri.nih.gov
Class: shotgun.

FEATURES
source
Location/Qualifiers
1..895
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="260L13"
/clone_lib="CITB"

ORIGIN
Query Match 89.5%; Score 17; DB 28; Length 895;
Best Local Similarity 100.0%; Pred. No. 2e+03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CCTTCGCGAAAAAATC 18
|||

Db 662 CCTTCGCGAAAAAATC 678
|||

RESULT 5
CC277174/c

LOCUS
DEFINITION 1161 bp DNA linear GSS 13-MAY-2003
CH261-22G8 RM1.1 CH261 Gallus gallus genomic clone CH261-22G8,
genomic survey sequence.

ACCESSION
VERSION CC277174

KEYWORDS
SOURCE GSS.

ORGANISM
Gallus gallus (chicken)

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.

TITLE
JOURNAL 1 (bases 1 to 1161)

COMMENT Krenitzki, C., Higginbotham, J., Wylie, K., Carter, J., McPherson, J.,
Warren, W., Graves, T., Wardis, E. and Wilson, R.
Gallus gallus BAC End Reads
Unpublished (2003)
Contact: Richard K. Wilson
Genome Sequencing Center

Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 182000 Std Error: 0.00
Seq primer: RM1 TACGACTCACTATAGGAGA
Class: BAC ends
High quality sequence start: 64
High quality sequence stop: 387.

FEATURES
source
Location/Qualifiers
1..1161
/organism="Gallus gallus"
/mol_type="genomic DNA"
/strain="Red Jungle Fowl"
/db_xref="taxon:9031"
/clone="CH261-22G8"
/sex="female"
/cell_line="UCD001, inbred 256"
/clone_lib="CH261"

ORIGIN
Query Match 89.5%; Score 17; DB 28; Length 1161;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCTTCGCGAAAAAAT 17
|||

Db 1016 GCCTTCGCGAAAAAAT 1000
|||

RESULT 6
AV098279

LOCUS
DEFINITION 179 bp mRNA linear EST 22-NOV-1999
AV098279 Mus musculus C57BL/6J ES cell Mus musculus cDNA clone
2410042F04, mRNA sequence.

ACCESSION
VERSION AV098279 GI:5245827

KEYWORDS
SOURCE EST.

ORGANISM
Mus musculus (house mouse)

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 179)
Carninci, P., Shibata, K., Ozawa, Y., Konno, H., Itoh, M., Aizawa, K.,
Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T.,
Hara, A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M.,
Kawai, J., Kikuchi, N., Kojima, Y., Matsuyama, T., Nitsuma, H., Oda, H.,
Owa, C., Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y.,
Sugahara, Y., Suzuki, H., Suzuki, H., Tateo, M., Tomaru, Y.,
Tomimaga, N., Watanabe, S., Yagame, M., Yamamura, T., Yokota, T.,
Yoshino, M., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.

TITLE
JOURNAL Unpublished (1999)

COMMENT RIKEN Mouse ESTs
Contact: Chie Owa
Genome Science Laboratory
RIKEN
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-res@rtc.riken.go.jp
Thermotabilization and thermoactivation of thermostable enzymes by
trehalose and its application for the synthesis of full length cDNA
(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (<http://genome.rtc.riken.go.jp>) for
further details.

FEATURES
source
Location/Qualifiers
1..179
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"

```

/db_xref="taxon:10090"
/clone="2410042F04"
/cell_type="ES cell"
/clone_lib="Mus musculus C57BL/6J ES cell"

ORIGIN
Query Match      86.3%; Score 16.4; DB 9; Length 179;
Best Local Similarity 94.4%; Pred. No. 4.8e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCTTCGCGAAAAAATC 18
    |||||
Db 26 GCCTTCGCGAAAAAATC 43

RESULT 7
A1177223
LOCUS
DEFINITION
EST220834 Normalized rat ovary, Bento Soares Rattus sp. cDNA clone
ROVCD11 3' end, mRNA sequence.
ACCESSION
A1177223
VERSION
A1177223.1 GI:3727861
KEYWORDS
EST.
SOURCE
Rattus sp.
ORGANISM
Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 412)
AUTHORS
Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
Kerlavage,A.R. and Adams,M.D.
TITLE
Rat Genome Project: Generation of a Rat EST (RST) Catalog & Rat
Gene Index
JOURNAL
Unpublished (1998)
COMMENT
Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.

FEATURES
Location/Qualifiers
1..412
/mol_type="mRNA"
/db_xref="ATCC (inhost):2032932"
/db_xref="taxon:10118"
/clone="ROVCD11"
/clone_lib="Normalized rat ovary, Bento Soares"
/notes="Organ: ovary; Vector: pT73Pac; Site_1: EcoRI;
Site_2: NotI"

ORIGIN
Query Match      86.3%; Score 16.4; DB 9; Length 412;
Best Local Similarity 94.4%; Pred. No. 4.2e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 CCTTCGCGAAAAAATCG 19
    |||||
Db 197 CCTTCGCGAAGAAATCG 214

RESULT 8
BQ511872/c
LOCUS
DEFINITION
EST619287 Generation of a set of potato cDNA clones for microarray
analyses mixed potato tissues Solanum tuberosum cDNA clone STMH57
5' end, mRNA sequence.
ACCESSION
BQ511872
VERSION
BQ511872.2 GI:21927507
KEYWORDS
EST.
SOURCE
Solanum tuberosum (potato)
ORGANISM
Solanum tuberosum

Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Solanaceae; Solanum.
1 (bases 1 to 452)
Buell,C.R., Hart,A., Baker,B., Tanksley,S., Fry,W., Smart,C.,
Restrepo,S., Griffiths,H., van der Hoeven,R., Tsai,J. and
Karamycheva,S.A.
Generation of a set of potato cDNA clones for microarray analyses
Unpublished (2002)
On Jun 10, 2002 this sequence version replaced gi:21370741.
Other ESTs: EST619288
Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potato-array@tigr.org
This clone can be obtained from the University of Arizona Genomics
Institute. Orders can be made through URL:
http://genome.arizona.edu/orders/
Seq primer: T3.

FEATURES
Location/Qualifiers
1..452
/organism="Solanum tuberosum"
/mol_type="mRNA"
/cultivar="Kennebec or Binjite"
/db_xref="taxon:4113"
/clone="STMH57"
/tissue_type="mixed tissues"
/lab_hosts="SOLR"
/clone_lib="Generation of a set of potato cDNA clones for
microarray analyses mixed potato tissues"
/notes="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
XhoI; supplier: Combination of untreated and Phytophthora
infestans-treated libraries of stolons, leaves, leaflets,
axillary buds of stem explants, petioles, germinating
eyes, tubers, or roots."

ORIGIN
Query Match      86.3%; Score 16.4; DB 13; Length 452;
Best Local Similarity 94.4%; Pred. No. 4.1e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 CCTTCGCGAAAAAATCG 19
    |||||
Db 367 CCTTCGCGAAGAAATCG 350

RESULT 9
AZ049007
LOCUS
DEFINITION
GSBRU0407 Sheared genomic library Brucella melitensis biovar
Abortus genomic clone B3J15, genomic survey sequence.
ACCESSION
AZ049007
VERSION
AZ049007.1 GI:7272922
KEYWORDS
GSS.
SOURCE
Brucella melitensis biovar Abortus (Brucella abortus)
ORGANISM
Brucella melitensis biovar Abortus
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Brucellaceae; Brucella.
1 (bases 1 to 481)
Sanchez,D.O., Zandomeni,R., Cravero,S., Rosetti,O., Grau,O. and
Ugalde,R.
Gene discovery through genomic sequencing survey of the Brucella
abortus genome
Unpublished (1999)
Contact: Sanchez D.O.
Instituto de Investigaciones Biotecnologicas (Univ. Nac. de Gral
San Martin)
Av. Gral Paz S/N, INTI, Edificio 24, B 1650 KNA, San Martin, Buenos
Aires, Argentina
Tel: (54-11) 4580/7255/7
Fax: (54-11) 4752-9639
Email: dsanchez@iib.unsam.edu.ar
Seq primer: T7
```

```

Class: shotgun.
FEATURES
  source
    Location/Qualifiers
      1..481
        /organism="Brucella melitensis biovar Abortus"
        /mol_type="genomic DNA"
        /strain="S-2308"
        /db_xref="taxon:235"
        /clone="B3J15"
        /clone_lib="Sheared genomic library"
        /note="Vector: pBluescript SK(-) (STRATAGENE); Genomic DNA
        was mechanically sheared, blunt ended, and
        size-fractionated by agarose gel electrophoresis.
        Fragments between 1.5-3 Kb were recovered and ligated to
        the EcoRV site of the pBluescript SK (-) vector."

ORIGIN
Query Match      86.3%; Score 16.4; DB 28; Length 481;
Best Local Similarity 94.4%; Pred. No. 4.1e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCTTCGCGAAAAAATC 18
    |||||
Db 124 GCCTTCGCGAAAAAATC 141

RESULT 10
LOCUS BE343391/c
DEFINITION EST408553 potato stolon, Cornell University Solanum tuberosum cDNA
clone CSTA24G20, mRNA sequence.
ACCESSION BE343391
VERSION BE343391.1 GI:9252923
KEYWORDS EST.
SOURCE Solanum tuberosum (potato)
ORGANISM Solanum tuberosum

REFERENCE
  1 (bases 1 to 531)
AUTHORS van der Hoeven,R., Bezzerides,J., Bachem,C., Horvath,B., Visser,R.,
Holt,I.E., Liang,F., Hansen,T.S., Usterback,T., Bowman,C.L., and
Doan,B., Bougri,O., Buell,C.R., Rohning,C.M., Tanksley,S.D. and
Baker,B.
Generation of ESTs from potato swelling stolons
Unpublished (1999)
Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potato-array@tigr.org
This clone can be obtained from the University of Arizona Genomics
Institute. Orders can be made through URL:
http://genome.arizona.edu/orders/.

FEATURES
  source
    Location/Qualifiers
      1..531
        /organism="Solanum tuberosum"
        /mol_type="mRNA"
        /cultiivar="Bintje"
        /db_xref="taxon:4113"
        /clone="CSTA24G20"
        /tissue_type="axillary buds of stem explants, swelling
        stolons"
        /dev_stages="1 to 3 days"
        /lab_host="SOUR"
        /clone_lib="potato stolon, Cornell University"
        /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
        XhoI; RNA was supplied by Christian Bachem & Beatrix_
        Horvath(Laboratory of Plant Breeding, Dept. of Plant
        Sciences, Wageningen University, The Netherlands). Total
        RNA was isolated from developing axillary buds of potato
        nodal stem cuttings cultured on medium for the
        introduction of tuber formation as described in Bachem et
        al. (Plant Journal 1996). Tissue samples were taken of
        stages corresponding to growing stolons and the early

```

```

ORIGIN
  source
    stages of tuber formation."
Query Match      86.3%; Score 16.4; DB 10; Length 531;
Best Local Similarity 94.4%; Pred. No. 4e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCTTCGCGAAAAAATCG 19
    |||||
Db 457 CCTTCGCGAAAAAATCG 440

RESULT 11
LOCUS BF496774
DEFINITION AT10684.5prime AT Drosophila melanogaster adult testes pOTB7
Drosophila melanogaster cDNA clone AT10684 5 similar to CG12736;
FBan0012736 located on: 2R 43D1-43D1:: 04/08/2001, mRNA sequence.
ACCESSION BF496774
VERSION BF496774.2 GI:13694252
KEYWORDS EST.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster

REFERENCE
  1 (bases 1 to 551)
AUTHORS Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Baxter,E.,
Bernan,B., Carlson,J., Champe,M., Chavez,C., Chew,M., Dorsett,V.,
Farfan,D., Frise,E., George,R., Gonzalez,M., Guarin,H., Harris,N.,
Li,P., Liao,G., Miranda,A., Miera,S., Mungall,C.J., Nunoo,J.,
Pacleb,J., Paragas,V., Park,S., Phouanavong,S., Wan,K., Yu,C.,
Lewis,S.E., Celniker,S. and Rubin,G.M.
BDGP/HMI AT Drosophila EST Project
Unpublished (2000)
On Dec 6, 2000 this sequence version replaced gi:11580075.
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
hit genomic AB003841: arm:2R [2297353,2559434]
estimated-cyto:43A1-43D3: 04/08/2001
Plate: AT.106 row: G column: 12
High quality sequence stop: 491.
Location/Qualifiers
  1..551
    /organism="Drosophila melanogaster"
    /mol_type="mRNA"
    /db_xref="taxon:7227"
    /clone="AT10684"
    /sex="male"
    /dev_stages="0-3 day old Ore-R males"
    /lab_host="Plates AT.10-AT.120: DHS-alpha. Plates
    AT.121-AT.319: DHS-alpha Tona"
    /clone_lib="AT Drosophila melanogaster adult testes pOTB7"
    /note="Organ: ADULT testes; Vector: pOTB7; Site 1: EcoRI;
    Site 2: XhoI. The mRNA for the testis library was made
    from testes and seminal vesicles hand dissected from 0-3
    day old Ore-R males. RNA kindly provided by the lab of
    Margaret Fuller. Sized fractionated cDNAs were directly
    ligated into pOTB7. Plasmid cDNA library."

ORIGIN
  source
    stages of tuber formation."
Query Match      86.3%; Score 16.4; DB 10; Length 551;
Best Local Similarity 94.4%; Pred. No. 4e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCTTCGCGAAAAAATCG 19
    |||||
Db 433 CCTTCGCGAAAAAATCG 450

```



```

RESULT 12
LOCUS      CC356196
DEFINITION PUHQZ77TB_ZM_0.6_1.0_KB_Zea_mays_genomic_clone_ZMMBTA514M10,
            599 bp DNA linear GSS 16-MAY-2003
ACCESSION  CC356196
VERSION     CC356196
KEYWORDS    GSS.
SOURCE      Zea mays
ORGANISM    Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 599)
AUTHORS     Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
            Bennetzen,J.
            Maize Genomics Consortium
            Unpublished (2003)
            Other GSSs: PUHQZ77TD
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TR
            Class: sheared ends.
FEATURES    Location/Qualifiers
             1..599
                /organism="Zea mays"
                /mol_type="genomic DNA"
                /strain="B73"
                /db_xref="caxon:4577"
                /clone_lib="ZMMBTA514M10"
                /note="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
                Cot selected genomic DNA library"
ORIGIN
Query Match      86.3%; Score 16.4; DB 28; Length 599;
Best Local Similarity 94.4%; Pred. No. 3.9e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 CCTCGCGAAAAAATCG 19
Db      517 CATTCCGAAAAAATCG 534

FEATURES    source
LOCUS      CK242884
DEFINITION CK242884 688 bp mRNA linear EST 12-DEC-2003
            Solanum tuberosum cDNA clone POCAL59 5' end, mRNA sequence.
ACCESSION  CK242884
VERSION     CK242884
KEYWORDS    EST.
SOURCE      Solanum tuberosum (potato)
ORGANISM    Solanum tuberosum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; lamids; Solanales; Solanaceae; Solanum.
REFERENCE   1 (bases 1 to 688)
AUTHORS     Buell,C.R., Hart,A., Zismann,V., Karamycheva,S.A. and Baker,B.
            Generation of ESTs from potato callus tissue
            Unpublished (2003)
            Other ESTs: EST726520
            Contact: Robin Buell
            The Institute for Genomic Research
            9712 Medical Center Dr, Rockville, MD 20850, USA
            Email: potato-array@tigr.org
            Clones can be requested from TIGR via potato@tigr.org

RESULT 13
LOCUS      CC356196
DEFINITION PUHQZ77TB_ZM_0.6_1.0_KB_Zea_mays_genomic_clone_ZMMBTA514M10,
            599 bp DNA linear GSS 16-MAY-2003
ACCESSION  CC356196
VERSION     CC356196
KEYWORDS    GSS.
SOURCE      Zea mays
ORGANISM    Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 599)
AUTHORS     Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
            Bennetzen,J.
            Maize Genomics Consortium
            Unpublished (2003)
            Other GSSs: PUHQZ77TD
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TR
            Class: sheared ends.
FEATURES    Location/Qualifiers
             1..599
                /organism="Zea mays"
                /mol_type="genomic DNA"
                /strain="B73"
                /db_xref="caxon:4577"
                /clone_lib="ZMMBTA514M10"
                /note="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
                Cot selected genomic DNA library"
ORIGIN
Query Match      86.3%; Score 16.4; DB 28; Length 599;
Best Local Similarity 94.4%; Pred. No. 3.9e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 CCTCGCGAAAAAATCG 19
Db      517 CATTCCGAAAAAATCG 534

FEATURES    source
LOCUS      CC685856/c
DEFINITION OGVEJ52TV_ZM_0.7_1.5_KB_Zea_mays_genomic_clone_ZMMBMA0536108,
            774 bp DNA linear GSS 19-JUN-2003
            genomic survey sequence.
ACCESSION  CC685856
VERSION     CC685856
KEYWORDS    GSS.
SOURCE      Zea mays
ORGANISM    Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 774)
AUTHORS     Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
            Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
            Consortium for Malze Genomics
            Unpublished (2002)
            Other GSSs: OGVEJ52TH
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES    Location/Qualifiers
             1..774
                /organism="Zea mays"
                /mol_type="genomic DNA"
                /strain="B73"
                /db_xref="taxon:4577"
                /clone_lib="ZMMBMA0536108"
                /clone_lib="ZM_0.7_1.5_KB"
                /note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
                methylation filtered genomic DNA library"
ORIGIN
Query Match      86.3%; Score 16.4; DB 29; Length 774;
Best Local Similarity 94.4%; Pred. No. 3.8e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 CCTTCGCGAAAAAATCG 19
Db      355 CATTCCGAAAAAATCG 338

FEATURES    source
LOCUS      CC685856/c
DEFINITION OGVEJ52TV_ZM_0.7_1.5_KB_Zea_mays_genomic_clone_ZMMBMA0536108,
            774 bp DNA linear GSS 19-JUN-2003
            genomic survey sequence.
ACCESSION  CC685856
VERSION     CC685856
KEYWORDS    GSS.
SOURCE      Zea mays
ORGANISM    Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 774)
AUTHORS     Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
            Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
            Consortium for Malze Genomics
            Unpublished (2002)
            Other GSSs: OGVEJ52TH
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES    Location/Qualifiers
             1..774
                /organism="Zea mays"
                /mol_type="genomic DNA"
                /strain="B73"
                /db_xref="taxon:4577"
                /clone_lib="ZMMBMA0536108"
                /clone_lib="ZM_0.7_1.5_KB"
                /note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
                methylation filtered genomic DNA library"
ORIGIN
Query Match      86.3%; Score 16.4; DB 29; Length 774;
Best Local Similarity 94.4%; Pred. No. 3.8e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 CCTTCGCGAAAAAATCG 19
Db      355 CATTCCGAAAAAATCG 338

```

RESULT 15
 CK242883/c
 LOCUS CK242883 799 bp mRNA linear EST 12-DEC-2003
 DEFINITION EST726520 potato callus cdna library, normalized and full-length
 Solanum tuberosum cdna clone POCA159 5' end, mRNA sequence.
 ACCESSION CK242883
 VERSION CK242883.1 GI:39786948
 KEYWORDS EST.
 SOURCE Solanum tuberosum (potato)
 ORGANISM Solanum tuberosum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 asterids; lamids; Solanales; Solanaceae; Solanum.
 REFERENCE 1 (bases 1 to 799)
 AUTHORS Buell, C.R., Hart, A., Zisemann, V., Karamycheva, S.A. and Baker, B.
 TITLE Generation of ESTs from potato callus tissue
 JOURNAL Unpublished (2003)
 COMMENT Other ESTs: EST726521
 Contact: Robin Buell
 The Institute for Genomic Research
 9712 Medical Center Dr, Rockville, MD 20850, USA
 Email: potato-array@tigr.org
 Clones can be requested from TIGR via potato@tigr.org
 Seq primer: ATT TAG GTG ACA CTA TAG.
 FEATURES
 source
 1..799
 location/Qualifiers
 /organism="Solanum tuberosum"
 /mol_type="mRNA"
 /cultivar="Kennebec"
 /db_xref="taxon:4113"
 /clone="POCA159"
 /tissue_type="callus"
 /lab_host="DH10B-TorA"
 /clone_lib="potato callus cdna library, normalized and
 full-length"
 /note="Vector: pCMVSPORT6.1; Site 1: EcoRI; Site 2: NotI;
 supplier: RNA was isolated from Solanum tuberosum var.
 Kennebec callus tissue grown on solid media."

ORIGIN

Query Match 86.3%; Score 16.4; DB 14; Length 799;
 Best Local Similarity 94.4%; Pred. No. 3.8e+03;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 CCTTCGCGAAAAAATCG 19
 |||||
 DB 475 CCTTCGCGAGAAATCG 458
 |||||

Search completed: April 24, 2004, 21:10:59
 Job time : 1482.95 secs